

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 29, 2003, 13:24:48 ; Search time 3691 Seconds  
(without alignments)  
10529.435 Million cell updates/sec

Title: US-10-053-410-3

Perfect score: 950

Sequence: 1 aaaaaaacccctcgcgat.....aaaaaaaaaaaaaaaaaaaaa 950

Scoring table: IDENTITY NUC  
Gapop 10\_0 , Gapext 1.0

Searched: 2889711 seqs, 20454813386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : GenEmbl.\*

- 1: gb\_ba.\*
- 2: gb\_htg.\*
- 3: gb\_in.\*
- 4: gb\_om.\*
- 5: gb\_ov.\*
- 6: gb\_pat.\*
- 7: gb\_ph.\*
- 8: gb\_pl.\*
- 9: gb\_pr.\*
- 10: gb\_ro.\*
- 11: gb\_sts.\*
- 12: gb\_sy.\*
- 13: gb\_un.\*
- 14: gb\_vi.\*
- 15: em\_ba.\*
- 16: em\_fun.\*
- 17: em\_hum.\*
- 18: em\_in.\*
- 19: em\_mu.\*
- 20: em\_om.\*
- 21: em\_or.\*
- 22: em\_ov.\*
- 23: em\_pat.\*
- 24: em\_ph.\*
- 25: em\_pl.\*
- 26: em\_ro.\*
- 27: em\_sts.\*
- 28: em\_un.\*
- 29: em\_vi.\*
- 30: em\_htg\_hum.\*
- 31: em\_htg\_inv.\*
- 32: em\_htg\_other.\*
- 33: em\_htg\_mus.\*
- 34: em\_htg\_pin.\*
- 35: em\_htg\_rod.\*
- 36: em\_htg\_mam.\*
- 37: em\_htg\_vrt.\*
- 38: em\_sv.\*
- 39: em\_htgo\_hum.\*
- 40: em\_htgo\_mus.\*
- 41: em\_htgo\_other.\*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	932	98.1	932	8	AF371278	AF371278 Zea mays
2	171.8	18.1	757	8	OS19KDGLO	X63990 O.sativa mr
3	171.8	18.1	825	8	RICMSF	L12252 Oryza sativ
4	170.2	17.9	1773	8	D50643	D50643 Oryza sativ
5	170.2	17.9	99538	2	AC113332	AC113332 Oryza sat
6	170.2	17.9	124132	2	AC130605	AC130605 Oryza sat
7	158.2	16.7	102842	8	AF497474	AF497474 Aegilops
8	156.8	16.5	120562	8	AV268139	AV268139 Hordeum v
9	91.2	9.6	1859	8	HVDNAHOR3	X84368 H.vulgare H
10	91.2	9.6	2296	8	BLYHOR3	D82941 Barley Hor3
11	91.2	9.6	2380	6	BD012675	BD012675 Method of
12	91.2	9.6	2434	6	BD012674	BD012674 Method of
13	80	8.4	2804	8	AY248704	AY248704 Aegilops
14	80	8.4	2809	8	ASU39229	U39229 Aegilops ta
15	80	8.4	3095	8	TAGLU1DG	X03041 Wheat Gene
16	79.8	8.4	101083	2	AC139010	AC139010 Homo sapi
17	78.4	8.3	6462	8	TAGLD12B	X12929 Triticum ae
18	78.2	8.2	537	8	AF226698	AF226698 Aegilops
19	78	8.2	1494	8	AY299518	AY299518 Thinopyru
20	77.6	8.2	553	8	TAE308970	AJ308970 Triticum
21	77.6	8.2	598	8	TAE308967	AJ308967 Triticum
22	76.4	8.0	153180	2	AP004683	AP004683 Oryza sat
23	75.8	8.0	1905	8	AF476960	AF476960 Aegilops
24	75.2	7.9	906	8	AY263345	AY263345 Thinopyru
25	75	7.9	1980	8	AY263343	AY263343 Thinopyru
26	75	7.9	1998	8	ACY306973	AJ306973 Aegilops
27	74.8	7.9	1494	8	AY264065	AY264065 Thinopyru
28	74.4	7.8	1962	8	ACY306974	AJ306974 Aegilops
29	73.8	7.8	1052	11	PM2H12B	AL684839 Penicilli
30	73.6	7.7	2330	8	TAE314785	AJ314785 Triticum
31	73.6	7.7	2409	8	AF513640	AF513640 Aegilops
32	73.6	7.7	2996	8	X61026	X61026 Wheat Glu-1
33	73.4	7.7	1971	8	AF476962	AF476962 Aegilops
34	73.4	7.7	2163	8	AY245797	AY245797 Triticum
35	72.4	7.6	186752	9	AC111200	AC111200 Homo sapi
36	72	7.6	2331	8	AF216869	AF216869 Triticum
37	72	7.6	2331	8	SCE314781	AJ314781 Secale ce
38	71.8	7.6	1827	8	AY260548	AY260548 Triticum
39	71.8	7.6	1827	8	AY260549	AY260549 Triticum
40	71.8	7.6	1881	8	AY174159	AY174159 Aegilops
41	70.4	7.4	2297	8	AY249141	AY249141 Triticum
42	70.4	7.4	2331	8	SCE314774	AJ314774 Secale ce
43	70.4	7.4	2331	8	SCE314775	AJ314775 Secale ce
44	70.4	7.4	2331	8	SCE314777	AJ314777 Secale ce
45	70.4	7.4	2331	8	TAE314783	AJ314783 Triticum

ALIGNMENTS

RESULT 1  
AF371278  
LOCUS  
DEFINITION Zea mays alpha globulin mRNA, complete cds.  
ACCESSION AF371278  
VERSION AF371278.1 GI:16305141  
KEYWORDS  
SOURCE Zea mays  
ORGANISM Zea mays  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD  
clade; Panicoideae; Andropogoneae; Zea.  
REFERENCE 1 (bases 1 to 932)  
AUTHORS Woo,Y.M., Hu,D.W., Larkins,B.A. and Jung,R.  
TITLE Genomics analysis of genes expressed in maize endosperm identifies





FEATURES	Location/Qualifiers	3 (bases 1 to 99538)
source	1. .1773	Hsing, Y.-I. C. and Chow, T.-Y.
	/organism="Oryza sativa (japonica cultivar-group)"	Direct Submission
	/mol_type="genomic DNA"	Submitted (29-MAR-2003) Institute of Botany, Academia Sinica, 128,
	/db_xref="taxon:39947"	Section 2, Yen-chu-yuan Road, Nankang, Taipei 11529, Taiwan
	/tissue_type="endosperm"	The nucleotide sequence of this BAC clone was generated by
	/dev_stage="immature"	combining Monanto and ASPG-Taiwan sequencing data.
	1009_ .1569	* NOTE: This is a 'working draft' sequence. It currently
CDS	/codon_start=1	* consists of 5 contigs. Gaps between the contigs
	/product="26 kDa globulin"	* are represented as runs of N. The order of the pieces
	/protein_id="BAA09308.1"	* is believed to be correct as given, however the sizes
	/db_xref="GI:1783206"	* of the gaps between them are based on estimates that have
	/translation="MASKVFFFAALMAAMVAISGLSESMRFRDQCOREVQSDP	* provided by the submitter.
	LDCROVLRLQTLRRRPPRPGALGLRMOCCOQLODVSEPCRAIRMRVSYE	* This sequence will be replaced
	ENSPMLEQWSSSEYVYGGESSEQYVGGSEEGYVGGQQQPGMTRVRLTRA	* by the finished sequence as soon as it is available and
	RQYAAQPSMCRVPPQCSIFAAGQ"	* the accession number will be preserved.
BASE COUNT	471 a 448 c 448 g 406 t	
ORIGIN		
Query Match	17.9%; Score 170.2; DB 8; Length 1773;	
Best Local Similarity	71.3%; Pred. No. 1.4e-17;	
Matches 239; Conservative	0; Mismatches 93; Indels 3; Gaps 1;	
QY	138 GCAGCGCTGTGCTCGCGCCCTGCTGGCGCTGCTCTGCAAGCGCAGGTCGAGCGG	197
Db	1033 GCAGCGCGCTCATGCGCGCCATGTTGGCCATCTCCGCGCGCAGTGCAGTGGAG	1092
QY	198 CAGAGCTCAGGACCTGTCAGTCTGTCAGGAGGTTCAGAGAGCCGCTCGAGCGTGC	257
Db	1093 ATGAGGTTACAGGACAGCAGTCCAGCGCGGAGTGCAGACAGCCGCTGAGCGGTGC	1152
QY	258 CGCAGGCTCTCCAGCGGAGCTAACCGCGCGCGCGCGCGCGCGCTTGCCTTC	317
Db	1153 CGCAGGCTCTCCAGCGGAGCTAACCGCGCGCGCGCGCGCGCGCTTGCCTTC	1212
QY	318 CGGTGGCGGACCGCGCTCCGATGCTGTCAGCAGCTCCAGGACGTGAGCGCGAG	377
Db	1213 CGCGCGCGCTCGCGCTCGGATGTCAGTGTGTCAGCAGCTGTCAGGACCGCGAG	1272
QY	378 TGCCTGTCGCCCGCATCCGAGAGTGTGTCAGGGGTACAGAGGCGCATGCC---GCCG	434
Db	1273 TGCCTGTCGCCCGCATCCGCGGATGTGAGGAGTGTACAGGAGAGCATGCCGATGCC	1332
QY	435 CTGAGAGAGGCTGGTGGCCATCGGGCGCGCAGCA	469
Db	1333 CTGAGAGAGGCTGGTGGCTGCTGTCGTCGAGTA	1367
RESULT 5		
AC113332/c	99538 bp DNA linear HTG 29-MAR-2003	
LOCUS	Oryza sativa (japonica cultivar-group) chromosome 5 clone	
DEFINITION	OJ1057_B02, *** SEQUENCING IN PROGRESS ***	5 ordered pieces.
AC113332		
AC113332.1	GI:19033375	
VERSION	HTG; HTGS PHASE2.	
KEYWORDS	Oryza sativa (japonica cultivar-group)	
SOURCE	Oryza sativa (japonica cultivar-group)	
ORGANISM	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;	
	Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;	
	Ehrhartoidae; Oryzaceae; Oryza.	
REFERENCE	1 (bases 1 to 99538)	
AUTHORS	Chow, T.-Y., Hsing, Y.-I. C., Chen, C.-S., Chen, H.-H., Wu, H.-P.,	
	Liu, S.-M., Chao, Y.-T., Chang, S.-J., Chen, T.-R., Chen, Y.-L.,	
	Chow, M.-H. J., Hong, Y.-C., Hsiung, J.-N., Hsu, C.-H., Huang, J.-J.,	
	Kau, P.-I., Lee, M.-C., Leu, H.-L., Lin, S.-J., Wu, L.-F. and Shaw, J.-F.	
	Oryza sativa BAC OJ1057_B02 genomic sequence	
JOURNAL	Unpublished	
REFERENCE	2 (bases 1 to 99538)	
AUTHORS	Chow, T.-Y. and Hsing, Y.-I. C.	
TITLE	Direct Submission	
JOURNAL	Submitted (01-MAR-2002) Institute of Botany, Academia Sinica, 128,	
	Section 2, Academia Road, Nankang, Taipei 11529, Taiwan	
FEATURES	source	
	1. .99538	/organism="Oryza sativa (japonica cultivar-group)"
		/mol_type="genomic DNA"
		/cultivar="Nipponbare"
		/db_xref="taxon:39947"
		/chromosomes="5"
		/clone="OJ1057_B02"
		/note="japonica cultivar-group"
BASE COUNT	26954 a 22452 c 22701 g 27030 t	401 others
ORIGIN		
Query Match	17.9%; Score 170.2; DB 2; Length 99538;	
Best Local Similarity	71.3%; Pred. No. 1.3e-17;	
Matches 239; Conservative	0; Mismatches 93; Indels 3; Gaps 1;	
QY	138 GCGCGCTGTGCTCGCGCCCTGCTGGCGCTGCTGCGCGCGCAGTGCAGCGG	197
Db	34084 GCGCGCGCTCATGCGCGCCATGTTGGCCATCTCCGCGCGCAGTGCAGTGCAG	34025
QY	198 CAGAGCTCAGGACCTGTCAGTCTGTCAGGAGGTTCAGAGAGCCGCTCGAGCGTGC	257
Db	34024 ATGAGGTTTCCAGGACAGGAGTCCAGCGGAGGTTCAGGACAGCCCGCTGAGCGGTGC	33965
QY	258 CGCCAGGCTCTCGACCGGAGCTAACCGCGCGCGCGCGCGCGCTTGCCTTC	317
Db	33964 CGCAGGCTCTCGACCGGAGCTAACCGCGCGCGCGCGCGCGCTTGCCTTC	33905
QY	318 CGGTGGCGGACCGCGCTCCGATGCTGTCAGCAGCTCCAGGACGTGAGCGCGAG	377
Db	33904 CGCGCGCGCTCGCGCTCGGATGTCAGTGTGTCAGCAGCTGTCAGGACCGGAG	33845
QY	378 TGCCTGTCGCCCGCATCCGAGAGTGTGTCAGGGGTACAGAGGCGCATGCC---GCCG	434
Db	33844 TGCCTGTCGCCCGCATCCGCGGATGTGTCAGGAGTGTACAGGAGAGCATGCCGATGCC	33785
QY	435 CTGAGAGAGGCTGGTGGCCATCGGGCGCGCAGCA	469
Db	33784 CTGAGAGAGGCTGGTGGCTGCTGTCGTCGAGTA	33750
RESULT 6		
AC130605/c	124132 bp DNA linear HTG 13-AUG-2002	
LOCUS	Oryza sativa (japonica cultivar-group) chromosome 5 clone P0010D04,	
DEFINITION	*** SEQUENCING IN PROGRESS ***	4 ordered pieces.
AC130605		
AC130605.1	GI:221212956	
VERSION		



[illegible]





```
RESULT 9
HVDNAHOR3
LOCUS H.vulgare Hor3 gene. 1859 bp DNA linear PLN 25-APR-1996
DEFINITION X84368
ACCESSION X84368
VERSION X84368.1 GI:671536
KEYWORDS D hordein; Hor3 gene.
SOURCE Hordeum vulgare subsp. vulgare
ORGANISM Hordeum vulgare subsp. vulgare
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Poideae; Triticeae; Hordeum.
REFERENCE 1
AUTHORS Sorensen,M.B., Muller,M., Skerriitt,J. and Simpson,D.
TITLE Hordein promoter methylation and transcriptional activity in
wild-type and mutant barley endosperm
JOURNAL Mol. Gen. Genet. 250 (6), 750-760 (1996)
MEDLINE 96204516
PUBMED 8628236
REFERENCE 2 (bases 1 to 1859)
AUTHORS Sorensen,M.B.
TITLE Direct Submission
JOURNAL Submitted (01-FEB-1995) M.B. Sorensen, Carlsberg Laboratory,
Department of Physiology, Gamle Carlsbergvej 1, DK-2500 Valby,
DENMARK
COMMENT overlapping with that under the accession number X68072.
FEATURES
source
1. .1859
/organism="Hordeum vulgare subsp. vulgare"
/mol_type="genomic DNA"
/variety="Boni"
/db_xref="taxon:112509"
/chromosome="5"
/map="1H long arm"
/clone="phor3-1"
/tissue_type="endosperm"
/clone_lib="lambda Zap II"
/dev_stage="seed"
TATA_signal 343. .349
gene 435. .1859
CDS 435..>1859
/gene="Hor3"
/codon_start=1
/product="D-hordein"
/protein_id="CAA59104.1"
/db_xref="GI:671537"
/db_xref="SPTRMBL_040045"
/translations="MAKRLVLFVAVIVALTAREINGNNIFLDSRSLQOCERE
LQESSLEACRRVDDQLVGLPMTGLQCCQLRDVSPCRPVALSQVVRQYEQQT
EVPKSGSYFGTAPPLQGGWGTSVKMYPDQTSQQSWGQGGYHCVTSQQP
GCGQGGYFGSTFPQGGQGGPQGPQWPSATFPQGGQGGYFGATSLQ
PGQGGYFGSTFPQGGQGGPQGPQWPSATFPQGGQGGYFGATSLQ
QSGQGGYFGSTFPQGGQGGPQGPQWPSATFPQGGQGGYFGATSLQ
SQGGQGGYFGSTFPQGGQGGPQGPQWPSATFPQGGQGGYFGATSLQ
HQQQLVGGQGGYFGSTFPQGGQGGPQGPQWPSATFPQGGQGGYFGATSLQ
SQGGVQGGYFGSTFPQGGQGGPQGPQWPSATFPQGGQGGYFGATSLQ"
mat_peptide 498. .>1859
/gene="Hor3"
/product="unnamed"
variation 1500
/gene="Hor3"
variation 1514
/gene="Hor3"
variation 1810
/gene="Hor3"
BASE COUNT 548 a 550 c 462 g 299 t
ORIGIN
Query Match 9.6%; Score 91.2; DB 8; Length 1859;
Best Local Similarity 61.4%; Pred. No. 5.1e-05;
```

```
Matches 178; Conservative 0; Mismatches 88; Indels 24; Gaps 1;
QY 207 AGGGACCTGCTGCTGGCAGGAGGTCCAGGAGAGCCCTCCAGCGTGCCTCCAGGTC 266
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 543 AGGCAGCTACAGTGTGAGCGCGAGCTCCAGGAGAGTCTGCTCAGCGTGCCTCCAGGTC 602
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 267 CTGACCGCGCAGCTAAACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 326
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 603 GTGGACCAACAGCTG-----GTTGCCAGCTGCCATGAGC 638
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 327 ACCGGCTCCGATGCGGTGCTGCCAGAGTCCAGAGCTGAGCGAGTGCCTCCAGTGC 386
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 639 ACGGGGCTCCAGATGAGTGTGCTGCCAGAGCTTGGGAGCTCAGCGCGAGTGCCTCCAGTGC 698
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 387 GCGGCCATCGGAGCATGCTGAGGCTACGAGGAGCCATGCGCGCGCTGGAGAAAGC 446
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 699 GTGCGCTCAGCGAGTGTGAGGCAATACGAGCAGCAACCGAGTGCCTCCAGTGC 758
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 447 TGGTGCCATGCGGGCGCGCAGCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 496
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 759 GATCTCTTCTACCGCGCGGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 808
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
RESULT 10
BLVHOR3
LOCUS Barley Hor3 mRNA for D hordein, complete cds. 2296 bp mRNA linear PLN 06-FEB-1999
DEFINITION D82941
ACCESSION D82941.1 GI:1167497
VERSION D hordein; Hor3.
KEYWORDS Hordeum vulgare subsp. vulgare
SOURCE Hordeum vulgare subsp. vulgare
ORGANISM Hordeum vulgare subsp. vulgare
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Poideae; Triticeae; Hordeum.
REFERENCE 1 (bases 1 to 2296)
AUTHORS Hirota,N., Kuroda,H. and Ito,K.
JOURNAL Unpublished
AUTHORS 2 (bases 1 to 2296)
TITLE Hirota,N.
JOURNAL Direct Submission
TITLE Submitted (06-JAN-1996) Naohiko Hirota, Plant Bioengineering
Research Laboratories, Sapporo breweries, Biotechnology department;
Kizaki 37-1, Nitta, Gunma 370-03, Japan
(E-mail:sappant@po.infosphere.or.jp, Tel:0276-56-1455,
Fax:0276-56-1605)
FEATURES
source
1. .2296
/organism="Hordeum vulgare subsp. vulgare"
/mol_type="mRNA"
/strain="Haruna Nijo"
/db_xref="taxon:112509"
/clone="DH4"
/tissue_type="Endosperm"
/dev_stage="Seed"
1. .2296
/gene="Hor3"
37. .2160
/gene="Hor3"
/function="seed storage"
/codon_start=1
/product="D hordein"
/protein_id="BAAL1642.1"
/db_xref="GI:1167498"
translation="MAKRLVLFVAVIVALTAREINGNNIFLDSRSLQOCERE
LQESSLEACRRVDDQLVGLPMTGLQCCQLRDVSPCRPVALSQVVRQYEQQT
EVPKSGSYFGTAPPLQGGWGTSVKMYPDQTSQQSWGQGGYHCVTSQQP
GCGQGGYFGSTFPQGGQGGPQGPQWPSATFPQGGQGGYFGATSLQ
PGQGGYFGSTFPQGGQGGPQGPQWPSATFPQGGQGGYFGATSLQ
QSGQGGYFGSTFPQGGQGGPQGPQWPSATFPQGGQGGYFGATSLQ
SQGGQGGYFGSTFPQGGQGGPQGPQWPSATFPQGGQGGYFGATSLQ
HQQQLVGGQGGYFGSTFPQGGQGGPQGPQWPSATFPQGGQGGYFGATSLQ
SQGGVQGGYFGSTFPQGGQGGPQGPQWPSATFPQGGQGGYFGATSLQ"
gene
CDS
```







[illegible]



**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 29, 2003, 13:28:03 ; Search time 2415 Seconds  
(without alignments)  
9560.767 Million cell updates/sec

Title: US-10-053-410-3

Perfect score: 950

Sequence: 1 aaaaaaacccctcgtcat.....aaaaaaaaaaaaaaaaaaaaa 950

Scoring table: IDENTITY NUC

Gapop 10\_0 , Gapext 1.0

Searched: 22781392 seqs, 12152238056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:\*

1: em\_estba:\*  
2: em\_esthum:\*  
3: em\_estin:\*  
4: em\_estnu:\*  
5: em\_estov:\*  
6: em\_estpl:\*  
7: em\_estro:\*  
8: em\_htc:\*  
9: gb\_est1:\*  
10: gb\_est2:\*  
11: gb\_htc:\*  
12: gb\_est3:\*  
13: gb\_est4:\*  
14: gb\_est5:\*  
15: em\_estfun:\*  
16: em\_estom:\*  
17: em\_gss\_hum:\*  
18: em\_gss\_inv:\*  
19: em\_gss\_pin:\*  
20: em\_gss\_vrt:\*  
21: em\_gss\_fun:\*  
22: em\_gss\_mam:\*  
23: em\_gss\_mus:\*  
24: em\_gss\_pro:\*  
25: em\_gss\_rod:\*  
26: em\_gss\_phg:\*  
27: em\_gss\_vrl:\*  
28: gb\_gss1:\*  
29: gb\_gss2:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result	No.	Score	Match	Length	DB	ID	Description
C	1	894.2	94.1	935	29	CC349094	CC349094 OGUEN26TH
	2	893	94.0	900	11	AY105070	AY105070 Zea mays
	3	732	77.1	1038	29	CC349105	CC349105 OGUEN26TV
C	4	648.6	68.3	674	14	CA401604	CA401604 ELOIN0423

1040	C	5	520.2	54.8	570	14	CA402183	CA402183 ELOIN0432
1040	C	6	497	52.3	602	9	AI711821	AI711821 605064E02
1040	C	7	456.4	48.0	539	9	AI745997	AI745997 605078E03
1040	C	8	430.6	45.3	560	9	AI712201	AI712201 605065F05
1040	C	9	385.2	40.5	473	14	CD443922	CD443922 ELOIN0432
1040	C	10	374	39.4	586	10	BE917827	BE917827 OVI_7_F02
1040	C	11	362.6	38.2	605	10	BE917909	BE917909 OVI_7_F02
1040	C	12	340.2	35.8	367	12	BM500544	BM500544 PACO00000
1040	C	13	307.2	32.3	389	9	AI670620	AI670620 605035H03
1040	C	14	285	30.0	652	14	CA402267	CA402267 ELOIN0434
1040	C	15	256.2	27.0	425	10	EG048804	EG048804 OVI_23_CO
1040	C	16	209.8	22.1	496	10	EG049097	EG049097 OVI_23_CO
1040	C	17	170.2	17.9	443	9	AU182658	AU182658 AU182658
1040	C	18	170.2	17.9	453	9	AU182650	AU182650 AU182650
1040	C	19	170.2	17.9	522	9	AU076088	AU076088 AU076088
1040	C	20	170.2	17.9	537	13	BQ172833	BQ172833 1011b04_r
1040	C	21	170.2	17.9	595	13	BQ172799	BQ172799 1008b04_r
1040	C	22	168.8	17.8	444	14	C20089	C20089 C20089 Rice
1040	C	23	167.8	17.7	460	14	C20009	C20009 C20009 Rice
1040	C	24	167.8	17.7	461	14	C20289	C20289 C20289 Rice
1040	C	25	166.8	17.6	712	9	AU163963	AU163963 AU163963
1040	C	26	166.8	17.6	704	9	AU094496	AU094496 AU094496
1040	C	27	165.4	17.4	466	9	AU182628	AU182628 AU182628
1040	C	28	165.2	17.4	459	14	C20342	C20342 C20342 Rice
1040	C	29	164.4	17.3	452	14	C20476	C20476 C20476 Rice
1040	C	30	163.2	17.2	549	9	AA751601	AA751601 96AS0337
1040	C	31	163	17.2	687	9	AU163860	AU163860 AU163860
1040	C	32	161.8	17.0	683	9	AU164031	AU164031 AU164031
1040	C	33	161.6	17.0	707	9	AU164041	AU164041 AU164041
1040	C	34	161.4	17.0	497	13	BQ247039	BQ247039 TaE15002B
1040	C	35	161.4	17.0	534	14	CA734375	CA734375 wde2f.pk0
1040	C	36	161.4	17.0	536	13	BQ246513	BQ246513 TaE15009D
1040	C	37	161.4	17.0	599	13	BQ246084	BQ246084 TaE15016E
1040	C	38	161.4	17.0	600	13	BQ246182	BQ246182 TaE15015B
1040	C	39	161.4	17.0	603	13	BQ246465	BQ246465 TaE15010D
1040	C	40	161.4	17.0	619	13	BQ241153	BQ241153 TaE05008E
1040	C	41	161.4	17.0	625	13	BQ806663	BQ806663 WHE3581.G
1040	C	42	160.6	16.9	476	14	CA734371	CA734371 wde2f.pk0
1040	C	43	159.8	16.8	547	9	AU165811	AU165811 AU165811
1040	C	44	159.6	16.8	393	13	BQ244740	BQ244740 TaE15038B
1040	C	45	159.6	16.8	440	13	BQ246456	BQ246456 TaE15010E

## ALIGNMENTS

RESULT 1	CC349094/c	935 bp	DNA	linear	GSS 16-MAY-2003
LOCUS	OGUEN26TH ZM 0.7 1.5 KB	Zea mays	genomic clone	ZMMBMA0422E04,	
DEFINITION	genomic survey sequence.				
ACCESSION	CC349094				
VERSION	CC349094.1	GI:30818501			
KEYWORDS	GSS.				
SOURCE	Zea mays				
ORGANISM	Zea mays				
REFERENCE	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD clade; Panicoideae; Andropogoneae; Zea.				
AUTHORS	1 (bases 1 to 935) A., Fraser, C.M., Budiman, M.A., Bedell, J.A., Rohlfing, T., Citek, R.W., Nunberg, A., Robbins, D. and Lakey, N.				
TITLE	Consortium for Maize Genomics				
JOURNAL	Unpublished				
COMMENT	Contact: Cathy Whitelaw TIGR				
	9712 Medical Center Drive, Rockville, MD 20850, USA				
	Tel: 301-838-5843				
	Fax: 301-838-0208				
	Email: whitelaw@tigr.org				
	Seq primer: TR				
	Class: sheared ends.				

```
FEATURES
source
  Location/Qualifiers
    1..935
      /organism="Zea mays"
      /mol_type="genomic DNA"
      /strain="B73"
      /db_xref="taxon:4577"
      /clone="ZMMBMA0422E04"
      /clone_lib="ZM 0.7-1.5 KB"
      /notes="Vector: pBCSK-; Site 1: HincII; 0.7-1.5 kb
      methylation filtered genomic DNA library"
BASE COUNT    147 a    339 c    259 g    190 t
ORIGIN
Query Match      94.1%; Score 894.2; DB 29; Length 935;
Best Local Similarity 98.0%; Pred. No. 1.4e-80;
Matches 916; Conservative 0; Mismatches 18; Indels 1; Gaps 1;

QY 11 CCTCGTGCATCACCACCAAG-AACACAGTAACCTAGCAGCTAGCACATCAACAAGTGGC 69
DB 935 CCTCGTGCATCACCACCAAGAACACAGTAACCTAGCAGCTAGCACATCAACAAGTGGC 876
QY 70 GACAGACAAAGATTGTGAGGGTGATCCGCGCTGAGAGAGATGGCTAAGATCGCCGCG 129
DB 875 GACAGACAAAGATTGTGAGGGTGATCCGCGCTGAGAGAGATGGCTAAGATCGCCGCG 816
QY 130 CGCGCGCGCGCGCTGTGCTTCCGCGCCCTGTGCGCGCTGCGCGCTGCGCGCGAGG 189
DB 815 CGCGCGCGCGCGCTGTGCTTCCGCGCCCTGTGCGCGCTGCGCGCTGCGCGCGAGG 756
QY 190 TCAGCGCGCAGAGGCTCAGGAGCTCGCAGTGTGCGCAGAGAGTCCAGGAGAGCCGCTCG 249
DB 755 TCAGCGCGCAGAGGCTCAGGAGCTCGCAGTGTGCGCAGAGAGTCCAGGAGAGCCGCTCG 696
QY 250 ACGGTGTCGCCAGGTCTTCGACCGGAGTAAACCGCGCGCGCGCGCGCGCGCGGTG 309
DB 695 ACGGTGTCGCCAGGTCTTCGACCGGAGTAAACCGCGCGCGCGCGCGCGCGCGGTG 636
QY 310 GCCGTTCGGTGGGCGACCGGCTCCGATGCGGTGCTGCCAGCAGCTCCAGAGGTGA 369
DB 635 GCCGTTCGGTGGGCGACCGGCTCCGATGCGGTGCTGCCAGCAGCTCCAGAGGTGA 576
QY 370 GCCCGAGTCCGCTGCGCGCGCATCCGAGAGCTGTCAGGGGCTACGAGAGGCCATGC 429
DB 575 GCCCGAGTCCGCTGCGCGCGCATCCGAGAGCTGTCAGGGGCTACGAGAGGCCATGC 516
QY 430 CGCGCTGGAGAAAGCTGTGTGCCATGCGGGCGCGCAGCAGCGCGCGCGCGAG 489
DB 515 CGCGCTGGAGAAAGCTGTGTGCCATGCGGGCGCGCAGCAGCGCGCGCGCGAG 456
QY 490 GAGGAGGAGACAGGGGGGTACTACTACCCCTGCAGCGCGCGCAGAGAGGGATACGGCT 549
DB 455 GAGGAGGAGACAGGGGGGTACTACTACCCCTGCAGCGCGCGCAGAGAGGGATACGGCT 396
QY 550 ACGGTCAAGGTGCGCAGCGCAGATGTATCACCGTGTGCTCCGCGCACCCCGCGCGG 609
DB 395 ACGGTCAAGGTGCGCAGCGCAGATGTATCACCGTGTGCTCCGCGCACCCCGCGCGG 336
QY 610 GGCACAGGATCGCGCGGTGAGGCTTACGAAGGCCCGGGAGTACGCGCGGGGTTCGCGA 669
DB 335 GGCACAGGATCGCGCGGTGAGGCTTACGAAGGCCCGGGAGTACGCGCGGGGTTCGCGA 276
QY 670 TGATGTCGGCTGTGCGAGCCCCAGAGTGCAGCATCTTCTCCGGGGGACAGTACT 729
DB 275 TGATGTCGGCTGTGCGAGCCCCAGAGTGCAGCATCTTCTCCGGGGGACAGTACT 216
QY 730 AGCTACCATGTTAAAGCAGAGTCCGCGCAGGTGCAAGACGACGATGTGTACTGTGCGC 789
DB 215 AGCTACCATGTTAAAGCAGAGTCCGCGCAGGTGCAAGACGACGATGTGTACTGTGCGC 156
QY 790 GTCAAAATCCAGAAATGAGTGTCTGACGTGGGCTCGCAATATTGTGCGGTTCGTGTA 849
DB 155 GTCAAAATCCAGAAATGAGTGTCTGACGTGGGCTCGCAATATTGTGCGGTTCGTGTA 96

FEATURES
source
  Location/Qualifiers
    1..900
      /organism="Zea mays"
      /mol_type="mRNA"
      /db_xref="MaizeDB:635953"
      /db_xref="taxon:4577"
      /clone_lib="Maize Mapping Project/DuPont Cornsensus
      Library"
      /note="this sequence is part of a project of EST
      assemblies resulting from the application of public
      contigs to seed DuPont contigs; this resource was
      assembled by Dupont as part of a collaboration for the
      overgo addressing of BACs in conjunction with the Maize
      Mapping Project"
BASE COUNT    191 a    244 c    331 g    134 t
ORIGIN
Query Match      94.0%; Score 893; DB 11; Length 900;
Best Local Similarity 100.0%; Pred. No. 1.9e-80;
Matches 893; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 51 AGCATCATCAACAAGTGGCGCAGACAGAAAGATTGTGAGGGTGATCCGCGCTGAGAAGAG 110
DB 8 AGCATCATCAACAAGTGGCGCAGACAGAAAGATTGTGAGGGTGATCCGCGCTGAGAAGAG 67
QY 111 ATGGCTTAAGATCCGCGCGCGCGCGCGCTGTGCTTCGCGGCCCTGGTGGCGGTG 170
DB 68 ATGGCTTAAGATCCGCGCGCGCGCGCGCTGTGCTTCGCGGCCCTGGTGGCGGTG 127
QY 171 GCGGTCTGCCAAGGCGAGGTTCAGCGCGCAGAGGCTCAGGAGACCTGAGTGTGCGAGGAG 230
DB 128 GCGGTCTGCCAAGGCGAGGTTCAGCGCGCAGAGGCTCAGGAGACCTGAGTGTGCGAGGAG 187
QY 231 GTCCAGGAGAGCCGCTCGACGCGTGCCTCGCCAGGTCTCTCGCCGAGCTAACCCGCGGC 290
```

Db	188	GTCCAGAGAGCCCGCTCGACGCTGCCGCGAGTCTCTGACCGGAGCTAAACCGCGGC	247
Qy	291	GGCGGGGGGGGGCGGTTCGCGCTTCGCTGGGGCACCGGGCTCCGGATCGGTGCTGC	350
Db	248	GGCGGGGGGGGGCGGTTCGCGCTTCGCTGGGGCACCGGGCTCCGGATCGGTGCTGC	307
Qy	351	CAGCAGCTCCAGAGCCTGAGCGCGAGTCCGCTGCGCGCCATCCGAGCATGTGTCAGG	410
Db	308	CAGCAGCTCCAGAGCCTGAGCGCGAGTCCGCTGCGCGCCATCCGAGCATGTGTCAGG	367
Qy	411	GGCTACGAGAGCCATGCGCGCTTCGAGAAAGGCTGTGGCCATGGGGGGGGGAG	470
Db	368	GGCTACGAGAGCCATGCGCGCTTCGAGAAAGGCTGTGGCCATGGGGGGGGGAG	427
Qy	471	CAGCGCGCGCGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG	530
Db	428	CAGCGCGCGCGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG	487
Qy	531	CCAGGAGAGGATACGGCTACGCTACGGGTGGCCAGCGGAGATGATCCACCGTGTCT	590
Db	488	CCAGGAGAGGATACGGCTACGCTACGGGTGGCCAGCGGAGATGATCCACCGTGTCT	547
Qy	591	CCCGGACCAACCGCGCGGCGCAAGATCGCGCGGTGAGGCTTACGAAGCGCGGGAG	650
Db	548	CCCGGACCAACCGCGCGGCGCAAGATCGCGCGGTGAGGCTTACGAAGCGCGGGAG	607
Qy	651	TACCGCGCGGGGTTCGGATGATGCGCGCTGTCCGAGCCCGAGAGTGCAGCATCTTC	710
Db	608	TACCGCGCGGGGTTCGGATGATGCGCGCTGTCCGAGCCCGAGAGTGCAGCATCTTC	667
Qy	711	TCCGCGCGCGACAGTACTAGCTTACCATTGTTAAAGCGAGTCCGCGCGAGGTGCAAGCG	770
Db	668	TCCGCGCGCGACAGTACTAGCTTACCATTGTTAAAGCGAGTCCGCGCGAGGTGCAAGCG	727
Qy	771	CAGCATGTGTTACTGCGCGTCAAAATCCAGAAATGACGTAGCTCTGACGTGGCTCGCAA	830
Db	728	CAGCATGTGTTACTGCGCGTCAAAATCCAGAAATGACGTAGCTCTGACGTGGCTCGCAA	787
Qy	831	TATTGTCGGTGTTCGTTACAAATGATAATTAATCATGAGGATTAATATGGATGTT	890
Db	788	TATTGTCGGTGTTCGTTACAAATGATAATTAATCATGAGGATTAATATGGATGTT	847
Qy	891	GCAGATAGTACTGCGCGCGGTCTTCAAAAAAAGGAGGAGGAGGAGGAGGAGGAGGAG	943
Db	848	GCAGATAGTACTGCGCGCGGTCTTCAAAAAAAGGAGGAGGAGGAGGAGGAGGAGGAG	900
FEATURES	Seq primer: TF Class: sheared ends. Location/Qualifiers 1..1038 /organism="Zea mays" /mol_type="genomic DNA" /strain="B73" /db_xref="taxon:4577" /clone="ZMMBMA0422E04" /clone_lib="ZM 0.7 1.5 KB" /note="Vector: pBSK-; Site 1: HincII; 0.7-1.5 kb methylation filtered genomic DNA library"		
BASE COUNT	215 a	316 c	336 g 171 t
ORIGIN	Query Match 77.1%; Score 732; DB 29; Length 1038; Best Local Similarity 99.7%; Pred. No. 1.7e-64; Matches 754; Conservative 0; Mismatches 0; Indels 2; Gaps 2;		
Qy	1	AAAAAAAAACCCCTCGTCGATCAACCAAGAACACAGTAACCTAGCAGCTAGCACATCAA	60
Db	283	AAAAAAAAACCCCTCGTCGATCAACCAAGAACACAGTAACCTAGCAGCTAGCACATCAA	342
Qy	61	ACAAGTGGCGACAGACAAAGATTGTGAGGGTGATCCGCGCTGAGAAGAGATGGCTAAGA	120
Db	343	ACAAGTGGCGACAGACAAAGATTGTGAGGGTGATCCGCGCTGAGAAGAGATGGCTAAGA	402
Qy	121	TGCGCGCGCGCGCGCGCGCGCTGTCTTCGCGCGCCCTGCTGGCGCTGCGCTGCGC	180
Db	403	TGCGCGCGCGCGCGCGCGCGCTGTCTTCGCGCGCCCTGCTGGCGCTGCGCTGCGC	462
Qy	181	AAGGCGAGGTCTGAGCGCGAGAGGCTCAGGAGCTTCAGTGTCTGGCAGGAGGTCCAGAGA	240
Db	463	AAGGCGAGGTCTGAGCGCGAGAGGCTCAGGAGCTTCAGTGTCTGGCAGGAGGTCCAGAGA	522
Qy	241	GCCCGCTCGACCGCTGCGCGCGAGTCTTCGACCGCGAGCTAACCGCGCGCGCGCGCG	300
Db	523	GCCCGCTCGACCGCTGCGCGCGAGTCTTCGACCGCGAGCTAACCGCGCGCGCGCGCG	582
Qy	301	GGGGGTTGGCCCGTTCGGTGGGGCACCGGGCTCCGATGCGGTGCTGCCAGCAGCTCC	360
Db	583	GGGGGTTGGCCCGTTCGGTGGGGCACCGGGCTCCGATGCGGTGCTGCCAGCAGCTCC	642
Qy	361	AGGAGTGAAGCGCGAGTGCCTGCGCGCGCATCCGAGAGCATGTGTCAGGGGTACGAGG	420
Db	643	AGGAGTGAAGCGCGAGTGCCTGCGCGCGCATCCGAGAGCATGTGTCAGGGGTACGAGG	702
Qy	421	AGGCCATGCGCGCGCTGGAGAAAGGCTGGTGGCCATGGGGGGCGGACAGCAGCCCGCG	480
Db	703	AGGCCATGCGCGCGCTGGAGAAAGGCTGGTGGCCATGGGGGGCGGACAGCAGCCCGCG	762
Qy	481	CGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG	540
Db	763	CGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG	822
Qy	541	GATACGGCTACGCTCAGGGTGGCCAGCGGAGATGTATCCACCGTGTGCTCCGCGCACCA	600
Db	823	GATACGGCTACGCTCAGGGTGGCCAGCGGAGATGTATCCACCGTGTGCTCCGCGCACCA	882
Qy	601	CGCGCGCGCGGCAAGGATCGCGCGCTGAGGCTTACGAAGGCCCGGAGGTACGCGCGG	659
Db	883	CGCGCGCGCGGCAAGGATCGCGCGCTGAGGCTTACGAAGGCCCGGAGGTACGCGCGG	942
Qy	660	GGGTTGCCGATGATGTGCGGTGTGCGAGCCCCAGGAGTGCAGCATCTTCTCCGG- CGG	718
Db	943	GGGTTGCCGATGATGTGCGGTGTGCGAGCCCCAGGAGTGCAGCATCTTCTCCGGCGG	1002
Qy	719	CGACGAGTACTAGCTACCATGGTTAAAGCGAGTCCG 754	
Db	1003	CGACGAGTACTAGCTACCATGGTTAAAGCGAGTCCG 1038	
RESULT	4		

RESULT 3  
CC349105  
LOCUS  
DEFINITION  
CC349105  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD clade; Panicoideae; Andropogoneae; Zea.  
1 (bases 1 to 1038)  
White, C.A., Quackenbush, J., Van Aken, S., Utterback, T., Resnick, A., Fraser, C.M., Budiman, M.A., Bedell, J.A., Rohlfing, T., Citek, R.W., Nunberg, A., Robbins, D., and Lakey, N.  
Consortium for Maize Genomics  
Unpublished  
Contact: Cathy Whitelaw  
TIGR  
9712 Medical Center Drive, Rockville, MD 20850, USA  
Tel: 301-838-5843  
Fax: 301-838-0208  
Email: whitelaw@tigr.org









```
ACCESSION CD43922
VERSION: CD43922.1 GI:31359565
KEYWORDS: EST.
SOURCE: Zea mays
ORGANISM: Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
clade; Panicoideae; Andropogoneae; Zea.
REFERENCE 1 (bases 1 to 473)
AUTHORS: Lai, J., Dey, N., Kim, C.S., Becraft, P., Larkins, B., Linton, E. and
Messing, J.
Sequencing of the maize endosperm ESTs
Unpublished
Contact: Lai, Jinsheng
Dr. Joachim Messing's lab
Waksman Institute, Rutgers University
190 Frelinghuysen Rd., Piscataway, NJ 08854, USA
Tel: 732-445-3801
Fax: 732-445-5735
Email: jlai@waksman.rutgers.edu
Seq primer: T3.
FEATURES             Location/Qualifiers
     source           1..473
                     /organism="Zea mays"
                     /mol_type="mRNA"
                     /cultivar="W22"
                     /db_xref="taxon:4577"
                     /tissue_type="Endosperm of 7-23DAP"
                     /clone_lib="Endosperm_4"
                     /notes="Vector: pBluescript SK-; Site_1: EcoRI; Site_2:
                     XhoI"
BASE COUNT   76 a 138 c 201 g 58 t
ORIGIN
Query Match      40.5%; Score 385.2; DB 14; Length 473;
Best Local Similarity 95.4%; Pred. No. 8.3e-30;
Matches 433; Conservative 0; Mismatches 8; Indels 13; Gaps 3;
QY 75 ACAAGATTGTCAGGTTGATCC-GCGCTCAGAGAGATGGCTAAGATCGCC---CGGCG 130
Db 29 ACAAGATTGTCAGAGTGATCCGGCGCTCAGAGAGATGGCTAAGATCGCCCGGCGGC 88
QY 131 GCGCGCGCGCGCGCTGTGCTTCGCGCCCTGGTGGCGCTGGCGCTCTGCCAAGCGAGGT 190
Db 89 GCGCGCGCGCGCGCTGTGCTTCGCGCCCTGGTGGCGCTGGCGCTCTGCCAAGCGAGGT 148
QY 191 CGAGCGCGCAGAGCTCAGGACCTGCAGTCTGGCAGAGGTCCAGAGAGCCCGCTCGA 250
Db 149 CGAGCGCGCAGAGCTCAGGCGCGGGAGTGTCTGGCAGAGGGCCAGGAGAGCCCGCTCGA 208
QY 251 CGGTGCGCGCAGTCTCTCGACCGGAGCTAACCGGCGCGCGCGCGCGCGCGCTGG 310
Db 209 CGGTGCGCGCAGTCTCTCGACCGGAGCTAACCGGCGCGCGCGCGCGCGCGCTGG 259
QY 311 CCCTTTCCGTGGGGCACCACCGGCTCGCGATCGCGTGTCCAGCAGCTCCAGGACGTGAG 370
Db 260 CCCTTTCCGTGGGGCACCACCGGCTCCGATCGCGTGTCCAGCAGCTCCAGGACGTGAG 319
QY 371 CCCTGAGTCCCGTGGCGCCCATCCGAGCATGGTTCAGGGGCTACAGAGAGCCCATGCC 430
Db 320 CCCTGAGTCCCGTGGCGCCCATCCGAGCATGGTTCAGGGGCTACAGAGAGCCCATGCC 379
QY 431 GCGCTGGAGAAAGGTGTGGCCATGGGGCGGCGCAGCAGCCCGCGCGCGAGG 490
Db 380 GCCTCTGGAGAAAGGTGTGGCCATGGGGCGGCGCAGCAGCCCGCGCGAGG 490
QY 491 AGGAGGAGGACAGGGGGGCTTACTACTACCCCTGC 524
Db 440 AGGAGGAGGACAGGGGGGCTTACTACTACCCCTGC 473
RESULT 10
BE917827
```

```
LOCUS BE917827
DEFINITION OV1.7_F02_g1_A002 Ovary 1 (OV1) Sorghum bicolor cDNA, mRNA
SEQUENCE.
ACCESSION BE917827
VERSION BE917827.1 GI:10420322
KEYWORDS Sorghum bicolor (sorghum)
SOURCE Sorghum bicolor
ORGANISM Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
clade; Panicoideae; Andropogoneae; Sorghum.
REFERENCE 1 (bases 1 to 586)
AUTHORS: Cordonnier-Pratt, M.-M., Gingle, A., Marsala, C., Sudman, M. and Pratt,
L.H.
An EST database from Sorghum: ovaries of varying immature stages
Unpublished
Contact: Cordonnier-Pratt, MM
Laboratory for Genomics and Bioinformatics
The University of Georgia, Department of Plant Biology
Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
Tel: 706 542 1860
Fax: 706 583 0210
Email: mmpratt@uga.edu
Sequences have been trimmed to exclude PolyA, vector and regions
below Phred quality 16. The threshold for highest quality sequence
is 20.
Seq primer: PolyTWix
High quality sequence start: 3
High quality sequence stop: 584
POLYA=No.
FEATURES             Location/Qualifiers
     source           1..586
                     /organism="Sorghum bicolor"
                     /mol_type="mRNA"
                     /db_xref="taxon:4558"
                     /clone_lib="Ovary 1 (OV1)"
                     /note="Organ: Mix of ovaries of varying immature stages
                     from 8-week-old plants; Vector: pBluescript II from Lambda
                     Zap II; Site 1: XhoI; Site 2: EcoRI; The library was made
                     from poly-A RNA in the cloning vector lambda ZAP II.
                     Clones to be sequenced were prepared by mass excision."
BASE COUNT   115 a 179 c 212 g 80 t
ORIGIN
Query Match      39.4%; Score 374; DB 10; Length 586;
Best Local Similarity 83.8%; Pred. No. 9.3e-29;
Matches 503; Conservative 0; Mismatches 45; Indels 52; Gaps 5;
QY 173 CGTCTGCCAAGCGAGGTCGAGCGGCAGAGGCTCAGGGACCTGCGTGTGGCAGGAGGT 232
Db 1 CGTCGCGCCAGGCGGTGTCGAGCGGCAGAGGCTCAGAGACCTGCTGCTGGCAGAGGT 60
QY 233 CCAGGAGAGCCCGCTCGACGCGTGGCGCGCAGGTCCTCGACCGGCGAGCTAACCGGCGCGG 292
Db 61 CCAGGAGAGACCCGCTCGGCGCGTGGCGCGCAGGTCGTCGACCGGCGAGCTCAC---CGGCGG 117
QY 293 CGGCGCGCGCGCGCTTGGCCCGTTCGCTGGGCGCACCGGCTCCGATCGGCTGCTGCCA 352
Db 118 CATGCGCTACGGCATCGGCCCGTTCGCTGGGCGCACCGGCTCCGATCGGCTGCTGCCA 177
QY 353 GCAGCTCCAGGAGCTGAGCGCGGAGTGGCGCTGGCGCGCATCCGAGAGATGTCAGGGG 412
Db 178 GCAGCTCCAGGAGCTGAGCGCGGAGTGGCGCTGGCGCGCATCCGAGAGATGTCAGGGG 237
QY 413 CTACGAGGAGGCGCATGCCCGCTCGAGAAAGCTGCTGGTGGCCATGGGGCGGCGGAGCA 472
Db 238 CTACGAGGAGAGCCATGCCCGCTCGAGAAAGCTGCTGGTGGCCATGGGGCGGCGGAGCA 291
QY 473 GCGGCGCGCGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 532
Db 292 GCC-----GGGCTACGACTACCTTACCTTACCTTACCTTACCTTACCTTACCTTACCTT 321
QY 533 AGGAGAGGATACGGGTACGGTTCAGGGTGGCCAGCGGAGATGTATCCACCGTGTCTGTC 592
```





Db 275 AAAAAAAAAACCTCATCGATCACCAACAAAGAACACAGTAACCTAGCTAGCTAGCAGATCAA 334  
Qy 61 ACAAGTGGGACAG-----ACAAGATTGTGAGGGTGATCC-CGGCTGAGAAGAGATGGC 115  
Db 335 ACAAGTGGGACAGCTCTACAAGATTGTGAGAGTGATCCGGCTGTGAGAAGAGATGGC 394  
Qy 116 TAAGATCGCCCGCGCGCGCGCGC---GGCGCTGTGCTTCGGCGCCCTGGTGCCCGTGGC 172  
Db 395 TAAGATCGCCCGCGCGCGCGCGCGCGCTGTGCTTCGGCGCCCTGGTGCCCGTGGC 454  
Qy 173 CGTCTGCCAAGGCGAGGTGCAGCGCGCAGAGCGCTCAGGACCTCAGTGTGCGCAGAGGT 232  
Db 455 CGTCTGCCAAGGCGAGGTGCAGCGCGCAGAGCGCTCAGGACCTCAGTGTGCGCAGAGGT 514  
Qy 233 CAGGAGAGCGCGCTGCAGCGCTCGCGCGCAGGTCTCGACCGCGCAGCTAACCGCGCGCG 292  
Db 515 CCAGGAGAGCGCGCTGCAGCGCTCGCGCGCAGGTCTCGACCGCGCAGCTAAC----- 565  
Qy 293 CGCGCGCGCGCGCTGGCGCGCTTCGGGTGGGCGCACCGCGCTCCGGATGCGGTGCGCA 352  
Db 566 CGGTGGCGCGCGCTGGCGCGCTTCGGCGCGCGCACCGCGCTCCGGATGCGGTGCGTGA 625  
Qy 353 GCAGCTCCAGGACGTGAGCGCGAGTG 379  
Db 626 GCAGCTCCAGGACGTGAGCGCGAGTG 652

## RESULT 15

BG048804  
LOCUS OVI\_23\_C05\_b1\_A002 Ovary 1 (Ov1) Sorghum bicolor cDNA, mRNA  
DEFINITION 425 bp mRNA linear EST 25-JAN-2001  
sequence.  
ACCESSION BG048804  
VERSION BG048804.1 GI:12499929  
KEYWORDS EST.  
SOURCE Sorghum bicolor (sorghum)  
ORGANISM Sorghum bicolor  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD  
clade; Panicoideae; Andropogoneae; Sorghum.  
REFERENCE 1 (bases 1 to 425)  
AUTHORS Cordonnier-Pratt,M.-M., Gingle,A., Marsala,C., Sudman,M. and Pratt  
L.H.  
TITLE An EST database from Sorghum: ovaries of varying immature stages  
JOURNAL Unpublished  
COMMENT Contact: Cordonnier-Pratt MM  
Laboratory for Genomics and Bioinformatics  
The University of Georgia, Department of Plant Biology  
Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA  
Tel: 706 542 1860  
Fax: 706 583 0210  
Email: mmpratt@uga.edu

Sequences have been trimmed to exclude PolyA, vector and regions  
below Phred quality 16. The threshold for highest quality sequence  
is 20.

Seq primer: JEN REV

High quality sequence stop: 393

POLYA=No.

## FEATURES

source

Location/Qualifiers

1..425  
/organism="Sorghum bicolor"  
/mol\_type="mRNA"  
/db\_xref="taxon:4558"  
/clone\_lib="Ovary 1 (Ov1)"  
/notes="Organ: Mix of ovaries of varying immature stages  
from 8-week-old plants; Vector: pBluescript II from Lambda  
Zap II; Site 1: XhoI; Site 2: EcoRI; The library was made  
from poly-A RNA in the cloning vector lambda ZAP II.  
Clones to be sequenced were prepared by mass excision."

BASE COUNT

ORIGIN

Query Match:

27.0%; Score 256.2; DB 10; Length 425;

Best Local Similarity 82.5%; Pred. No. 5.8e-17;  
Matches 362; Conservative 0; Mismatches 28; Indels 49; Gaps 4;  
Qy 334 TCAGATGCGGTGCTGCGCAGCAGCTCCAGGACGTGAGCCGCGAGTGCCCTGCGCCGCCA 393  
Db 1 TCAGATGCGGTGCTGCGCAGCAGCTCCAGGACGTGAGCCGCGAGTGCCCTGCGCCGCCA 60  
Qy 394 TCAGGAGCATGCTCAGGGGCTACGAGGAGCCATGCCCGCTGGAGAAAGCTGGTGGC 453  
Db 61 TCAGGAGCATGCTCAGGGGCTACGAGGAGACCATGCCCGCTGGAGAAAGCTGG- 116  
Qy 454 CATGGGGGCGGCGAGCAGCAGCCCGCCCGCAGGAGGAGGACAGAGGGGGGTACT 513  
Db 117 --TGGGGGCGAGCGCCAGCC-----GGGCTACG 144  
Qy 514 ACTACCCCTGCAGCGCGCCAGGAGGGATACGGCTACGGTCAGGGTGCCCGAGCGCAGA 573  
Db 145 ACTACCCCTGCAGCGCGCGAGGAGGGATACGGCTACGGTCAGAGTGCGCAGCAGA 204  
Qy 574 TGTATCCACCGTGTGCTCCCGGCACACCGCGCGGGGCCAAGGATCGGCCGCTGAGGC 633  
Db 205 TGTATCCACCGTGTGCTCCCGGCAC-----CGGGCAAAAAATCGGCCGCTGAAGC 255  
Qy 634 TTACGAAGGCCCGGAGTACGCCCGGGGTTCGCGATGATGTCCCGGCTGTTCGAGCCCC 693  
Db 256 TTACGAAGGCCCGGAGTACGCCCGGAGGATGCCCGATGATGTCCCGGCTGTTCGAGCCCC 315  
Qy 694 AGGAGTGACAGATCTTTCTCCCGCGCGCAGCAGTACTAGTACCATGTTAAAGCGAGTCG 753  
Db 316 AGGAGTGACAGCTCTTCTCCCGCGCGCAGCAGTA-----CTACTAGGGTAAGAGTGAGTCG 371  
Qy 754 GCGCGAGGTGCAAGAGCGCA 772  
Db 372 GCGCGCGGTGCAAGAGCGCA 390

Search completed: November 29, 2003, 15:52:57

Job time : 2422 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: November 29, 2003, 17:07:40 ; Search time 3510 Seconds  
(without alignments)  
2400.964 Million cell updates/sec

Title: US-10-053-410-4  
Perfect score: 1148  
Sequence: 1 MAKIAAAAAALCPAALVAV.....MMCRILSEPQECIFSGGDQY 206

Scoring table: BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 2888711 seqs, 2045481386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:  
-MODEL=frame+ p2n.model -DRV=xlh

-Q=/cgn2\_1/USFTO.spool/US10053410/runat\_28112003.140756.27880/app\_query.fasta\_1.391  
-DB=GenEmbl -QFMT=fastap -SUFFIX=p2n.rge -MINMATCH=0.1 -LOOPECL=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45  
-DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=pco -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=us10053410 @CGN 1.1 3508 @runat\_28112003.140756.27880 -NCPU=6 -ICPU=3  
-NO\_MMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

GenEmbl.\*  
1: gb.ba.\*  
2: gb.htg.\*  
3: gb.in.\*  
4: gb.om.\*  
5: gb.ov.\*  
6: gb.pat.\*  
7: gb.ph.\*  
8: gb.pl.\*  
9: gb.pr.\*  
10: gb.ro.\*  
11: gb.sts.\*  
12: gb.sy.\*  
13: gb.un.\*  
14: gb.vi.\*  
15: em.ba.\*  
16: em.fun.\*  
17: em.hum.\*  
18: em.in.\*  
19: em.mu.\*  
20: em.om.\*  
21: em.or.\*  
22: em.ov.\*  
23: em.pat.\*  
24: em.ph.\*  
25: em.pl.\*  
26: em.ro.\*  
27: em.sts.\*  
28: em.un.\*

29: em.vi.\*  
30: em.htg.hum.\*  
31: em.htg.inv.\*  
32: em.htg.other.\*  
33: em.htg.mus.\*  
34: em.htg.pln.\*  
35: em.htg.rod.\*  
36: em.htg.mam.\*  
37: em.htg.vrt.\*  
38: em.sv.\*  
39: em.hgo.hum.\*  
40: em.hgo.mus.\*  
41: em.hgo.other.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	1148	100.0	932	8 AF371278	AF371278 Zea mays
2	439	38.2	1773	8 D50643	D50643 Oryza sativ
3	439	38.2	99538	2 AC113332	AC113332 Oryza sat
4	439	38.2	124132	2 AC130605	AC130605 Oryza sat
5	435	37.9	757	8 OS19KDGL	X6390 O.sativa mr
6	433.5	37.8	825	8 RICMSF	LI252 Oryza sativ
7	414	36.1	102842	8 AF497474	AF497474 Aegilops
8	381	33.2	120562	8 AY268139	AY268139 Hordeum v
9	292.5	25.5	1962	8 ACY306974	AJ306974 Aegilops
10	289.5	25.2	3095	8 TAGLUIDG	X03041 Wheat gene
11	289	25.2	2163	8 AY245797	AY245797 Triticum
12	289	25.2	2330	8 TAE314785	AJ314785 Triticum
13	287	25.0	1980	8 AY263343	AY263343 Thinopyru
14	287	25.0	2996	8 X61026	X61026 Wheat Glu-1
15	286	24.9	906	8 AY263345	AY263345 Thinopyru
16	286	24.9	2220	8 AY263346	AY263346 Thinopyru
17	285	24.8	1494	8 AY299518	AY299518 Thinopyru
18	283	24.7	2409	8 AF513640	AF513640 Aegilops
19	282.5	24.6	1859	8 HVDNAHOR3	X84368 H.vulgare H
20	282	24.6	1494	8 AY264065	AY264065 Thinopyru
21	282	24.6	1788	8 AY298724	AY298724 Thinopyru
22	282	24.6	2809	8 ASU39229	U39229 Aegilops ta
23	281.5	24.5	2296	8 BLYHOR3	D82941 Barley Hor3
24	281.5	24.5	2380	6 BD012675	BD012675 Method of
25	281.5	24.5	2434	6 BD012674	BD012674 Method of
26	281	24.5	6462	8 TAGLD12B	X12929 Triticum ae
27	279	24.3	2804	8 AY248704	AY248704 Aegilops
28	277.5	24.2	598	8 TAE308967	AJ308967 Triticum
29	276.5	24.1	1371	8 AF476962	AF476962 Aegilops
30	275	24.0	2331	8 SCE314781	AJ314781 Secale ce
31	273.5	23.8	1881	8 AY174159	AY174159 Aegilops
32	271	23.6	1842	8 AY263344	AY263344 Thinopyru
33	269	23.4	2297	8 AY249141	AY249141 Triticum
34	267.5	23.3	2403	8 SCE314770	AJ314770 Secale ce
35	267	23.3	1920	8 AY245579	AJ245579 Triticum
36	267	23.3	1998	8 ACY306973	AJ306973 Aegilops
37	267	23.3	2313	8 SCE314767	AJ314767 Secale ce
38	267	23.3	2331	8 AF216869	AF216869 Triticum
39	267	23.3	2331	8 SCE314775	AJ314775 Secale ce
40	267	23.3	2331	8 SCE314777	AJ314777 Secale ce
41	267	23.3	2331	8 TAE314783	AJ314783 Triticum
42	264.5	23.0	537	8 AF226698	AF226698 Aegilops
43	262.5	22.9	1905	8 AF354289	AF354289 Aegilops
44	260	22.6	1905	8 AF476960	AF476960 Aegilops
45	259.5	22.6	1830	8 AY245578	AY245578 Triticum

ALIGNMENTS

RESULT 1



LDACROVLDRQTLGTRFRFQMPFRFPALGLRMCCQOQLODVSRRCRAAIRMRWSYE  
ESMPLEQMGSSSEYGGSESGYGGSEGGYGGQOQQPQMTWRVRLTRA  
ROYAALPQMCVRPEQCCSIFAAGQY"  
BASE COUNT 471 a 448 c 448 g 406 t  
ORIGIN

Alignment Scores:  
Pred. No.: 1,03e-14 Length: 1773  
Score: 439.00 Matches: 111  
Percent Similarity: 61.61% Conservatives: 19  
Best Local Similarity: 52.61% Mismatches: 52  
Query Match: 38.24% Indels: 30  
DB: 8 Gaps: 9

US-10-053-410-4 (1-206) x D50643 (1-1773)

QY 1 MetAlaLysIleAlaAlaAlaAlaAlaLeuCysPheAlaAlaLeuValAlaVal 20  
|||||  
Db 1009 ATGGCTAGCAAGTCTCTTCTTCGGCGCGCGCTC---ATGGCGGCCATGGTGGCCATC 1065  
QY 21 AlaValCysGlnGlyGluValGluArgGlnArgLeuArgAspLeuGlnCysTrpGlnGlu 40  
::: |||||  
Db 1066 TCCGGCGCGCAGTGCAGCGAGTCCGAGATCAGGTTTCAGGACAGGAGTCCAGCGGGAG 1125  
QY 41 ValGlnGluSerProLeuAspAlaCysArgGlnValLeuAspArgGlnLeuThrGly-Gl 60  
|||||  
Db 1126 GTGACGAGCAGCCGCTGACGCGTCCGCGAGTCTCCAGCGAGTCTCCAGCGCGCG 1185  
QY 60 YGlyGlyGlyGlyValGlyProPheArgTrpGlyThrGlyLeuArgMetArgCysCy 80  
|||||  
Db 1186 GAGAGTTCACGCCGAGTTCGCCGCCCGCGGC-GCGCTCGGCTCGGATGCGAGTCTG 1244  
QY 80 sGlnGlnLeuGlnAspValSerArgGluCysArgCysAlaAlaIleArgSerMetValAr 100  
|||||  
Db 1245 CCAGCAGCTGCAGGAGTGCAGCGCGAGTCCGCGCGCCTCCGCGCGAGTGTGAG 1304  
QY 100 gGlyTrpGluGluAlaMetPro---ProLeuGluLysGlyTrpTrpProTrpGlyArgGl 119  
::: |||||  
Db 1305 GAGCTACGAGGAGAGCATCCGATGCCCTGGAGCAAGGCTGCTCGTCTGTCGTCGCGA 1364  
QY 119 nGlnGlnProProGlnGlyGlyGlyGlyGlyGlnGlyGlyTrpTrpProCysSe 139  
: |||||  
Db 1365 GTACTAC-----GCGCGCGAGGGG-----TC 1385  
QY 139 rArgProGlyGluGlyTrpGlyTrpGlyGlnGlyGlnArgGlnMetTyr----- 156  
: |||||  
Db 1386 GTCTCGGAGGAGGGTAC---TACGCGAGGGTCTCGAGAGGGCTACTACGCGCA 1442  
QY 157 -ProProCysArgProGlyThrThrGlyGlyGlyProArgIleGlyArgValArgLeuTh 176  
::: |||||  
Db 1443 GCAGCAGCAGCAGCGCGGATGACC-----CGCGTGAGGCTGAC 1481  
QY 176 rLysAlaArgGluTyrAlaAlaGlyLeuProMetMetCysArgLeuSerGluProGlnGl 196  
::: |||||  
Db 1482 CAGGCGGAGCAGTACGCGCGCAGTCCGCGTGCATGTGCGGGTT---GAGCCCCAGCA 1538  
QY 196 uCysSerIlePheSerGlyGlyAspGlnTyr 206  
::: |||||  
Db 1539 GTGCAGCACTTCGCCCGCGC---CAGTAC 1566

## RESULT 3

AC113332/c  
LOCUS  
DEFINITION Oryza sativa (japonica cultivar-group) chromosome 5 clone  
OJ1057\_B02, \*\*\* SEQUENCING IN PROGRESS \*\*\*, 5 ordered pieces.  
AC113332  
AC113332.1 GI:19033375  
KEYWORDS HTG; HTGS PHASE2.  
SOURCE Oryza sativa (japonica cultivar-group)  
ORGANISM Oryza sativa (japonica cultivar-group)  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
Ehrhartoideae; Oryzaceae; Oryza.

## REFERENCE

AUTHORS  
1 (bases 1 to 99538)  
Chow, T.-Y., Hsiao, Y.-I. C., Chen, C.-S., Chen, H.-H., Wu, H.-P.,  
Liu, S.-M., Chao, Y.-T., Chang, S.-J., Chen, T.-R., Chen, Y.-L.,  
Chow, M.-H. J., Hong, Y.-C., Hsiung, J.-N., Hsu, C.-H., Huang, J.-J.,  
Kau, P.-I., Lee, M.-C., Leu, H.-L., Lin, S.-J., Wu, L.-F. and Shaw, J.-P.  
Oryza sativa BAC OJ1057\_B02 genomic sequence  
Unpublished

## JOURNAL

REFERENCE  
2 (bases 1 to 99538)

AUTHORS  
Chow, T.-Y. and Hsiao, Y.-I. C.

TITLE  
Direct Submission

JOURNAL  
Submitted (01-MAR-2002) Institute of Botany, Academia Sinica, 128,  
Section 2, Academia Road, Nankang, Taipei 11529, Taiwan

REFERENCE  
3 (bases 1 to 99538)

AUTHORS  
Hsiao, Y.-I. C. and Chow, T.-Y.

TITLE  
Direct Submission

JOURNAL  
Submitted (29-MAR-2003) Institute of Botany, Academia Sinica, 128,  
Section 2, Yien-chu-yuan Road, Nankang, Taipei 11529, Taiwan

COMMENT  
The nucleotide sequence of this BAC clone was generated by  
combining Monsanto and ASPQC-Taiwan sequencing data.

\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 5 contigs. Gaps between the contigs  
\* are represented as runs of N. The order of the pieces  
\* is believed to be correct as given, however the sizes  
\* of the gaps between them are based on estimates that have  
\* been provided by the submitter.

\* This sequence will be replaced  
\* by the finished sequence as soon as it is available and  
\* the accession number will be preserved.

\* 1 890: contig of 890 bp in length  
\* 891 990: gap of unknown length  
\* 991 19703: contig of 18713 bp in length  
\* 19704 19803: gap of unknown length  
\* 19804 75538: contig of 55735 bp in length  
\* 75539 75638: gap of unknown length  
\* 75639 92490: contig of 16852 bp in length  
\* 92491 92590: gap of unknown length  
\* 92591 99538: contig of 6948 bp in length.

FEATURES  
Location/Qualifiers

source  
1..99538

/organism="Oryza sativa (japonica cultivar-group)"  
/mol\_type="genomic DNA"

/cultivar="Nipponbare"  
/db\_xref="taxon:39947"

/chromosome="5"  
/clone="OJ1057\_B02"

/note="japonica cultivar-group"

BASE COUNT 26954 a 22452 c 22701 g 401 others

ORIGIN

Alignment Scores:

Pred. No.: 3.05e-13 Length: 99538

Score: 439.00 Matches: 111

Percent Similarity: 61.61% Conservatives: 19

Best Local Similarity: 52.61% Mismatches: 52

Query Match: 38.24% Indels: 30

DB: 2 Gaps: 9

US-10-053-410-4 (1-206) x AC113332 (1-99538)

QY 1 MetAlaLysIleAlaAlaAlaAlaAlaLeuCysPheAlaAlaLeuValAlaVal 20

Db 34108 ATGGCTAGCAAGTCTCTTCTTCGGCGCGCGCTC---ATGGCGGCCATGGTGGCCATC 34052

QY 21 AlaValCysGlnGlyGluValGluArgGlnArgLeuArgAspLeuGlnCysTrpGlnGlu 40

Db 34051 TCCGCGCGCAGCATGAGCGAGTCCGAGATGAGGTTTCAGGACAGGAGTCCAGCGGGAG 33992

QY 41 ValGlnGluSerProLeuAspAlaCysArgGlnValLeuAspArgGlnLeuThrGly-Gl 60

Db 33991 GTGCAGGACAGCCCGCTGACGCGTCCGCGAGGTCCTGACCGGAGCTCACCGCGCG 33932

QY 60 YGlyGlyGlyGlyValGlyValGlyProPheArgTrpGlyThrGlyLeuArgMetArgCysCy 80



```

Db 33931 GAGAGGTTCCAGCGCATGTTCCGCGCCCGCGGCGCTCGGCTCGGATGAGTGCCTG 33873
Qy 80 sGlnGlnLeuGlnAspValSerArgGluCysArgCysAlaAlaLeuValSerMetValar 100
Db 33872 CCAGCAGCTCAGGACGAGTGCAGCGCGAGTCCGCTCGCGCCCATCGCGGATGCTGAG 33813
Qy 100 gGlyTyrGluGluAlaMetPro---ProLeuGluLysGlyTyrTrpProTrpGlyArgG1 119
Db 33812 GAGCTACGAGGAGGATGCGATGCGCTCGGACAGGCTGCTGCTGCTGCTGCTGCTG 33753
Qy 119 nGlnGlnProProGlnGlnGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGly 139
Db 33752 GTACTAC-----GGCGCGAGGGG-----TC 33732
Qy 139 rArgProGlyGluGlyTyrGlyTyrGlyGlyGlyGlyGlyGlnArgGlnMetTyr----- 156
Db 33731 GTGCTCGGACAGGGGTAC---TACGGCGAGGGGTGCTCGGAGGAGGGCTACTACGGCGA 33675
Qy 157 -ProProCysArgProGlyThrThrGlyGlyGlyProArgIleGlyValArgValArgLeuTh 176
Db 33674 GCAGCAGCAGCAGCGGGGATGACC-----CGCGTGAAGCTGAC 33636
Qy 176 rLysAlaArgGluTyrAlaAlaGlyLeuProMetMetCysArgLeuSerGluProGlnG1 196
Db 33635 CAGCGGAGGAGCAGTACGCGCGCAGCTGCGCTCGATGTCGCGGGTT---GAGCCCCAGCA 33579
Qy 196 uCysSerIlePheSerGlyAspGlnTyr 206
Db 33578 GTGCAGCATCTTCGCGCGCGC---CAGTAC 33551

RESULT 4
AC130605/c 124132 bp DNA linear HTG 13-AUG-2002
LOCUS Oryza sativa (japonica cultivar-group) chromosome 5 clone P0010D04,
DEFINITION *** SEQUENCING IN PROGRESS ***, 4 ordered pieces.
ACCESSION AC130605
VERSION AC130605.1 GI:22212956
KEYWORDS HTG; HTGS PHASE2.
SOURCE Oryza sativa (japonica cultivar-group)
ORGANISM Oryza sativa (japonica cultivar-group)
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehrhartoideae; Oryzeae; Oryza.
1 (bases 1 to 124132)
Chow,T.-Y., Hsing,Y.-I.C., Chen,C.-S., Chen,H.-H., Liu,S.-M.,
Chao,Y.-T., Chang,S.-J., Chen,H.-C., Chen,S.-K., Chen,T.-R.,
Chen,Y.-L., Cheng,C.-H., Chung,C.-I., Han,S.-Y., Hsiao,S.-H.,
Hsiung,J.-N., Hsu,C.-H., Huang,J.-J., Kau,P.-I., Lee,M.-C.,
Leu,H.-L., Li,Y.-F., Lin,S.-J., Lin,Y.-C., Wu,S.-W., Yu,C.-Y.,
Yu,S.-W., Wu,H.-P. and Shaw,J.-F.
Oryza sativa PAC P0010D04 genomic sequence
Unpublished
TITLE Oryza sativa PAC P0010D04 genomic sequence
JOURNAL
REFERENCE 2 (bases 1 to 124132)
AUTHORS Chow,T.-Y. and Hsing,Y.-I.C.
Direct Submission
TITLE Oryza sativa (japonica cultivar-group)
JOURNAL
COMMENT
* NOTE: This is a 'working draft' sequence. It currently
* consists of 4 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as runs of N, however the sizes
* of the gaps between them are based on estimates that have
* been provided by the submitter.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
* 1 24826: contig of 24826 bp in length
* 24827 24926: gap of unknown length
* 24927 39423: contig of 14497 bp in length
* 39424 39523: gap of unknown length
* 39524 89605: contig of 50082 bp in length
* 89606 89705: gap of unknown length
* 89706 124132: contig of 34427 bp in length.

```

```

FEATURES
    source
        1. .124132
        /organism="Oryza sativa (japonica cultivar-group)"
        /mol_type="genomic DNA"
        /cultivar="Nipponbare"
        /db_xref="taxon:39947"
        /chromosome="5"
        /clone="P0010D04"
BASE COUNT 34211 a 27001 c 26916 g 35690 t 314 others
ORIGIN
Alignment Scores:
Pred. No.: 3,68e-13 Length: 124132
Score: 439.00 Matches: 111
Percent Similarity: 61.61% Conservative: 19
Best Local Similarity: 52.61% Mismatches: 52
Query Match: 38.24% Indels: 30
DB: 2 Gaps: 9
US-10-053-410-4 (1-206) x AC130605 (1-124132)
Qy 1 MetAlaLysIleAlaAlaAlaAlaAlaAlaLeuCysPheAlaAlaLeuValAlaVal 20
Db 119786 ATGGCTAGCAAGTCTCTTCTTCGCGCGCGGCTC---ATGCGCGCCATGTTGGCCATC 119730
Qy 21 AlaValCysGlnGlyGluValGlnArgGlnArgLeuArgAspLeuGlnCysTrpGlnGlu 40
Db 119729 TCCGCGCGCAGCTGAGCGAGTCCGAGATGAGTTCAGGACAGGCGAGTCCAGCGGAG 119670
Qy 41 ValGlnGlnSerProLeuAspAlaCysArgGlnValLeuAspArgGlnLeuThrGly-G1 60
Db 119669 GTGCAGGACAGCCCGTGGACGGTCCCGCAGGTGCTCAGCGCGCAGCTCACCAGCGCG 119610
Qy 60 yGlyGlyGlyGlyValGlyValGlyProPheArgTrpGlyThrGlyLeuArgMetArgCysCy 80
Db 119609 GAGAGTTCCAGCCGATGTTCCGCGCGCGGCGGCGCTCGCGCTCGCGATGCGAGTGC 119551
Qy 80 sGlnGlnLeuGlnAspValSerArgGluCysArgCysAlaAlaLeuArgSerMetValar 100
Db 119550 CCAGCAGCTGCGAGGACGTGAGCGCGAGTGCCTCGCGCGCCATCCGCGGATGCTGAG 119491
Qy 100 gGlyTyrGluGluAlaMetPro---ProLeuGluLysGlyTyrTrpProTrpGlyArgG1 119
Db 119490 GAGCTACGAGGAGGACATCCCGATGCTCCCTGGAGCAAGGCTGCTGCTGCTGCTGCG 119431
Qy 119 nGlnGlnProProGlnGlnGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGly 139
Db 119430 GTACTAC-----GGCGCGAGGGG-----TC 119410
Qy 139 rArgProGlyGluGlyTyrGlyTyrGlyGlnGlyGlnArgGlnMetTyr----- 156
Db 119409 GTGCTCGGAGCAGGGGTAC---TACGGCGAGGGGTGCTCGGAGGAGGGCTACTACGGCG 119353
Qy 157 -ProProCysArgProGlyThrThrGlyGlyGlyProArgIleGlyValArgValArgLeuTh 176
Db 119352 GCAGCAGCAGCAGCGGGGATGACC-----CGCGTGAAGCTGAC 119314
Qy 176 rLysAlaArgGluTyrAlaAlaGlyLeuProMetMetCysArgLeuSerGluProGlnG1 196
Db 119313 CAGGCGGAGGAGTACGCGCGCGAGTCCGCTCGATGTCGCGGTT---GAGCCCCAGCA 119257
Qy 196 uCysSerIlePheSerGlyAspGlnTyr 206
Db 119256 GTGCAGCATCTTCGCGCGCGC---CAGTAC 119229
RESULT 5
OS19KDGLO
LOCUS O. sativa mRNA for 19 kDa globulin.
DEFINITION X63990 S76663
ACCESSION X63990.1 GI:20158
VERSION 19 kDa globulin; alpha-globulin; cereal storage protein; globulin.
KEYWORDS Oryza sativa (japonica cultivar-group)
SOURCE

```

```

ORGANISM      Oryza sativa (japonica cultivar-group)
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehrhartoideae; Oryzeae; Oryza.
1 (bases 1 to 757)
Shorrosh,B.S., Wen,L., Zen,K.C., Huang,J.K., Pan,J.S.,
Hermanson,M.A., Tanaka,K., Muthukrishnan,S. and Reeck,G.R.
A novel cereal storage protein: molecular genetics of the 19 kDa
globulin of rice
Plant Mol. Biol. 18 (1), 151-154 (1992)
JOURNAL
MEDLINE 92119226
PUBMED 1731968
REFERENCE 2 (bases 1 to 757)
AUTHORS Shorrosh,B.
DIRECT SUBMISSION
TITLE Submitted (17-MAY-1992) B. Shorrosh, The Samuel Roberts Noble
Foundation, Plant Biology Division, 2510-1990East, Ardmore,
Oklahoma 73402, USA
FEATURES
source Location/Qualifiers
1..757
/organism="Oryza sativa (japonica cultivar-group)"
/mol_type="mRNA"
/strain="Nipponbare (japonica)"
/db_xref="taxon:19947"
/tissue_type="Endosperm"
/dev_stage="Endosperm, immature"
29..589
/codon_start=1
/product="19 kDa globulin precursor"
/protein_id="CAA45400.1"
/db_xref="GI:20159"
/db_xref="SWISS-PROT:P29835"
/translation="WASKVFFFAALMAAMVAISGAHVSEMRFRDROCOREVDSP
LDACQVLDRLTGTFERFQMRPCALGLRMCCQQLQDVRECECAIRWVRSYE
ESMPFLECGSSSEYEGSGSESGYEGEQQQPGQPPWRLVRLTRA
ROYAQLPSMCRVPEQQCSIFAAGQY"
29..94
sig_peptide
mat_peptide
95..586
/product="19 kDa globulin"
polya_signal
666..672
polya_signal
673..679
terminator
681..688
polya_signal
707..712
terminator
721..728
polya_signal
742
polya_signal
157 a 197 c 263 g 140 t
BASE COUNT 157 a 197 c 263 g 140 t
ORIGIN
Alignment Scores:
Pred. No.: 8-12e-15 Length: 757
Score: 435.00 Matches: 110
Percent Similarity: 61.14% Conservative: 19
Best Local Similarity: 52.13% Mismatches: 53
Query Match: 37.89% Indels: 30
DB: 8 Gaps: 9
US-10-053-410-4 (1-206) x OS19KDGLO (1-757)
Qy 1 MetAlaLysIleAlaAlaAlaAlaAlaLeuCysPheAlaAlaLeuValAlaVal 20
Db 29 ATGGCTAGCAAGGTCGCTCTTCGCGCGCGGCTC---ATGGCGGCCATGTGGGCATC 85
Qy 21 AlaValCysGlnGlyGluValGluArgGlnArgLeuAspLeuGlnCysTrpGlnGlu 40
Db 86 TCGGCGCGCAGCTGACGCGAGTCGCGAGATGAGGTCAGGACAGGCGAGTCCAGCGGAG 145
Qy 41 ValGlnGlnSerProLeuAspAlaCysArgGlnValLeuAspArgGlnLeuThrGly-Gl 60
Db 146 GTGCAGACAGACCCGCTGACGCGTCCGCGCAGTGTCTGACCGGAGCTACCGCGCGG 205
Qy 60 YGlyGlyGlyGlyGlyValGlyProPheArgTrpGlyThrGlyLeuArgMetArgCysCy 80
Db 206 GAGAGGTTCCAGCGGATGTTCCGCCCGCGGCGGCGGCTCGGCTCGGATGCGATGCTG 264

```

---

```

Qy 80 sGlnGlnLeuGlnAspValSerArgGluCysArgCysAlaAlaIleArgSerMetValAr 100
Db 265 CCAGCAGCTGCAGGACGTGAGCGCGGAGTGCCTGCGCGCCATCCGCGGATGCTGAG 324
Qy 100 gGlyTyrGluGluAlaMetPro---ProLeuGlnLysGlyTyrTrpProTrpGlyArgGl 119
Db 325 GAGCTACGAGGAGAGCATGCTCCGATGCCCTGAGCAAGGCTGCTCGCTCGCTCGCGGA 384
Qy 119 nGlnGlnProProGlnGlnGlyGlyGlyGlnGlnGlyGlyTyrTyrProCysSe 139
Db 385 GTACTAC-----GGCGCGGAGGGG-----TC 405
Qy 139 rArgProGlyGluGlyTyrGlyGlnGlnGlyGlnArgGlnMetTyr-----156
Db 406 GTCTGCGGAGCAGGGGTAC---TACGGCAGGGGTCTCGGAGGAGGCTACTACGGCGA 462
Qy 157 -ProProCysArgProGlyThrGlyGlyProArgGlyGlyArgValArgLeuTh 176
Db 463 GCAGCAGCAGCAGCGCGGATGACC-----CGCTGAGGCTGAC 501
Qy 176 rLysAlaArgGluTyrAlaAlaGlyLeuProMetMetCysArgLeuSerGluProGlnGl 196
Db 502 CAGGCGGAGCAGTACGCGCGCAGCTGCTGCGATGTCGCGGT---GAGCCCCAGCA 558
Qy 196 uCysSerIlePheSerGlyGlyAspGlnTyr 206
Db 559 GTGCAGCATCTTCGCGCGCGG---CAGTAC 586
RESULT 6
RICHMSF
LOCUS      RICHMSF
DEFINITION Oryza sativa DNA fragment with a miscellaneous signal and an open
ACCESSION L12252
VERSION L12252
KEYWORDS L12252.1 GI:169804
SOURCE      Oryza sativa (japonica cultivar-group)
ORGANISM    Oryza sativa (japonica cultivar-group)
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehrhartoideae; Oryzeae; Oryza.
Krishnan,H.B. and Pueppke,S.G.
Nucleotide sequence of an abundant rice seed globulin: homology
with the high molecular weight glutelins of wheat, rye and
triticales
Biochem. Biophys. Res. Commun. 193 (1), 460-466 (1993)
JOURNAL
MEDLINE 93277591
PUBMED 8503935
COMMENT Original source text: Oryza sativa (cultivar Lamont) 20 days
post-anthesis endosperm cDNA to mRNA.
FEATURES
source Location/Qualifiers
1..825
/organism="Oryza sativa (japonica cultivar-group)"
/mol_type="mRNA"
/cultivar="Lamont"
/db_xref="taxon:39947"
/tissue_type="endosperm"
/dev_stage="20 days post-anthesis"
1..54
misc_signal 25..549
CDS
/codon_start=1
/label=ORF
/protein_id="AAA72362.1"
/db_xref="GI:169805"
/translation="MAAMVAISGAHVSEMRFRDROCOREVDSPLDACROVLDROL
TGRERFQMRPFGALGLRMCCQQLQDVRECECAIRWVRSYEESMPFLEQMS
SSSRYEGGSGSSSQYEGESSEGYEGEQQQPGQPPWRLVRLTRARQYAAQLPSMCR
VEPQCSIFAAGQY"
165 a 244 c 270 g 146 t
BASE COUNT 165 a 244 c 270 g 146 t
ORIGIN

```





```

KNDISGTIPTSPAGLATVAVFSVARNHVHQVPPWLGNTLTALDLMADNMDSHGVPV
ALSKLINLRSLTVAINNLOGLIPVLFNMSLECLNFGSNQSGSLPQDQISGMLPNLK
KFGVFNREGQIPASLNSISLEHLSLHGNFRGRIPSNIGOSGLRTVFEVGNELQ
ATESRDWDLTSLANCSLLVNLQNLNSGLIPNSIGNLSQKLEGLRGVGNQIAGLI
PTGIGRYLKLAIEFADNRTGTIPSDIGKLSNKLKELSFQNFYIIEPSSIGNLSQL
NLIALSTNNLEGSIPATFNGNLTELI SLDIASNLLSGKIPFEEVNRISLALFNLSNNL
LDGPI SPHIGQLANLAI IDFSNKLGPPIPNALSGSCIALQFLHGNLQGGQIPKELM
ALRGLEBLDNLNLSGPVEFLESFOLLKNNLNFHLSGPPVDKGI FSNARVISLT
SNMGLQGGPVFPPTCPSPDKLASHKLLQILVTAVGARFILLGVCI AARCVNKS
RGDAHQQENIPMFQRTISYTELHSATDSFSEENLVGRSGFSGVSGSNGNLLTA
AVKVDVQROGATPRTFISECNALMIRHRKLKVITVCDLSHNGQFALVLEFIPN
DSMWLHGFSTEDFETPNMORNLALIDVAELEYLHDHIDPPIVHCDVKPSNILLD
GDMAHGFGLAKI IRAEKSQSLADQSCVSGIKGTIGYVAPEYGTGEISVEGDVY
SYGVLLLEMLTGRPTDPFSDTTLNPKVEMACPNLLETMDNIRCNQEPQAVLEL
FAAPVSRGLACGSGARQRIKMGDDVVKELGAIKQIIMASQNVASWSTKLY"
31751..31774
/repeat_region
/rpt_type=tandem
/rpt_unit=tctc
38555..38618
/transposon="MITE"
<38912..>39586
/product="globulin"
38912..39586
/codon_start=1
/product="globulin"
/db_xref="GI:30421166"
/translations="MKKFPFPAVFLTLVTISAAQVLQSADAOGRGEVQAKPLLA
QRIQLHQITGAVGVPPQAQGDRCQQLQESVRSRGCRCALRGMDVDEYQSMPP
LREGRSSRGREQEGESAEQVEQVGGYSGTESGQQQGGYGHVGTGVRG
GOOQGMCLRRPQROQGGEGFSGEQAQKPKVGRVRLTKVRLPTACRIEPOBCSVFS
TLPVLC"
complement(41245..42179)
/transposon="Solo-LTR retrotransposon Bare-1_184G9-2a"
complement(42180..42386)
/transposon="LTR retrotransposon Lollaog_184G9-1p"
complement(42387..43152)
/transposon="LTR retrotransposon Bare-1_184G9-3p"
complement(43153..53697)
/transposon="LTR retrotransposon Lollaog_184G9-1p"
complement(53698..53841)
/transposon="Solo-LTR retrotransposon Bare-1_184G9-2a"
53842..58439
/transposon="LTR retrotransposon Leojyg_184G9-1"
complement(58440..59149)
/transposon="Solo-LTR retrotransposon Bare-1_184G9-2a"
<62032..>64305
/product="D-Hordein"
62032..64305
/codon_start=1
/product="D-Hordein"
/db_xref="AAP31051.1"
/translations="MAKRLVLFVAVIVLVALVTTARREINGNNIFLDSRSLOQCERE
LOBSSLEACRVVDQOLQWPMSTGLQMCQCRDVSPECRPVALSQVRYEQEOT
EVPKSGSVFGGTAPLQOGWMGTSVKWYVDPQTSQSQWGQGHVQVTSQOP
GQGGQSYSGTFFPQPGGQCPQCPQSPYSATFPQPGQGGQGGYIPGATSLQ
PQGGQGGYVSATSPQPGGQGGQGGYPIATSPHQPGQGGQGGYIPGATSLQ
SQGGQGGYPTSPQPGGQGGQGGQGGQGGYPSATFPQPGQGGQGGYIPGATSLQ
SPQGGQGGQGGYPSATSPQPGQGGQGGQGGYPIATSPHQPGQGGQGGYIPGATSLQ
HGQLVQGGQGGQGGYPSATSPQPGQGGQGGYPIATSPHQPGQGGQGGYIPGATSLQ
SQGVQACQGHSTSPQQAQCAQSSPKQGLSLYPSGAYTOQKPGGYNPGTSP
LHQGGFGGGLTTPQGGQGGYPIATSPHQPGQGGYPIATSPHQPGQGGYIPGATSLQ
QTTVSPHQGGQGGYPIATSPHQPGQGGYPIATSPHQPGQGGYIPGATSLQ
QGGQGGQGGYPIATSPHQPGQGGYPIATSPHQPGQGGYIPGATSLQ
QGGQGGQGGYPIATSPHQPGQGGYPIATSPHQPGQGGYIPGATSLQ
LPMCRLEGGLLASQ"
join(<64782..64942,65077..65414,65452..65555,66035..66224,
66334..66725,66880..>67194)
/product="putative protein kinase"
join(<64782..64942,65077..65414,65452..65555,66035..66224,
66334..66725,66880..>67194)
/codon_start=1

```

```

/product="putative protein kinase"
/protein_id="AAP31052.1"
/db_xref="GI:30421168"
/translations="MIGTVLAHSDVDPAIRKVAARGRVVFAAETWGAARGRVVFA
AVRGGAARGHEABEPLTSETKKRKNVAASPDKKRMPKNIWKKKAKNQLATLV
GRGFCMAASHEABEPLTSETKKRKNVAASPDKKRMPKNIWKKKAKNQLATLV
KEISLATKNSAMHVNILNMIICPTQTYEHSYTLNLAIVIAVKQDKDGLQGNRE
FLVEVLMISLHHPNLVLLGYCTEQDKILVYEMPLGSLQDHLIDLTPKSOPLSMH
TMKTAADAARGLEYLHVANPPVVRDLKASNILLDGNFSAKLADFLAKLGPVGDK
THVTVWGTGYCYCAPEYAMSKLTKMSDIYCFGVVLELITGRRAIDTTPKPTREOIL
VHWAAPLFDKKMADPLDSDSKYPLKGLYQALASMSKLOEERASRLISDVVTA
LTFADPNYDPPDIEPLPISVPNDKGISLREABISLSGFBKQVEDS"
70339..77371
/transposon="non-LTR retrotransposon Leva_184G9-1"
70469..70482
/rpt_type=tandem
/rpt_unit=ta
complement(77378..>77855)
/transposon="LTR element Wobi_184G9-1p"
77856..86572
/transposon="LTR retrotransposon Bare-1_184G9-4"
complement(<86571..87579)
/transposon="LTR element Wobi_184G9-1p"
87266..87285
/rpt_type=tandem
/rpt_unit=ag
complement(87580..89305)
/transposon="LTR retrotransposon solo-LTR Bare-1_184G9-5a"
complement(<89306..90807)
/transposon="LTR element Wobi_184G9-1p"
complement(<91388..>91975)
/transposon="LTR retrotransposon Wham_184G9-1p"
complement(<93655..>95251)
/transposon="LTR retrotransposon Barbara_184G9-1p"
95284..>98782
/transposon="LTR retrotransposon Uaier_184G9-2p"
complement(<98783..100481)
/transposon="LTR retrotransposon Bare-1_184G9-6p"
100482..101938
/transposon="LTR retrotransposon solo-LTR Wham_184G9-2a"
101839..109132
/transposon="LTR element Egug_184G9-1"
109133..109156
/transposon="LTR retrotransposon solo-LTR Wham_184G9-2a"
complement(109157..>109388)
/transposon="LTR retrotransposon Bare-1_184G9-6p"
complement(109389..119184)
/transposon="LTR retrotransposon Wham_184G9-3"
complement(<119185..120562)
/transposon="LTR retrotransposon Bare-1_184G9-6p"
BASE COUNT 32679 a 27422 c 28340 g 32121 t
ORIGIN
Alignment Scores:
Pred. No.: 3-63e-10 Length: 120562
Score: 381.00 Matches: 98
Percent Similarity: 51.72% Conservative: 22
Best Local Similarity: 42.24% Mismatches: 50
Query Match: 33.19% Indels: 62
DB: 8 Gaps: 10
US-10-053-410-4 (1-206) x AY268139 (1-120562)
QY 11 AlaLeuCysePhaAlaLeuValAlaValAlaValCysGlnGlyCluValGluArgGln 30
Db 38933 GCGGTGTTCTCACGACCCCTGTGACCATCTCCGCCGCCCAAGGCGTCTCGAG---CAG 38989
QY 31 ArgLeuArgAspLeuGlnCysTrpGlnGluValGluSerProteinAspAlaCysArg 50
Db 38990 AGCCTCGCGGACGCCGAGTCCGGGGCGGAGGTCAGGCGAAGCCGCTCTCCGTCGCCG 39049
QY 51 GlnValLeuAspArgGlnLeuThrGlyGlyGlyGlyGlyGlyGlyValGlyProPheArg 70
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```





[illegible]



2 (bases 1 to 2163)  
Jiang, Y., Sun, M., Zheng, J., Xiao, Y. and Yan, Y.  
Direct Submission  
Submitted (03-MAR-2003) Key Lab of Genetics and Biotechnology,  
Biology Department, Capital Normal University, Beijing 100037, China

**FEATURES**

source	1..2163		
	/organism="Triticum turgidum subsp. durum"		
	/mol_type="genomic DNA"		
	/cultivar="Semito"		
	/sub_species="durum"		
	/db_xref="taxon:4567"		
gene	<1..>2163		
	/gene="Glu-1B"		
mRNA	<1..>2163		
	/gene="Glu-1B"		
	/product="HMW glutenin subunit 1By8"		
CDS	1..2163		
	/gene="Glu-1B"		
	/codon_start=1		
	/product="HMW glutenin subunit 1By8"		
	/protein_id="AAO64642.1"		
	/db_xref="GI:29150726"		
	/translation="MARLVLFATVITLVAAGEASRQLQCRELEQESSLEACRP QVVDTQAGRLPMSTGLQMRCSKRLNAVSVQRVEYETVVPKGGSFYP GETPTLQQLVFTWGTSVGYYPSVSSPQGQYPPQGAASPQPGQSQPEKWE LGCGQGYPTPSLSHQGGQYPPSSLPQPGQOIIGQCGGYPTSLLQPQSGQOI GQCGQGYPTSPHGPGRQPGQOGIIGQOLGGRROIIGQSGGQGGQGYPTSP QLCQGPQGWQSQGGQGYPTSCQPGQCGQGYPASQCPGQCGQSYPASQSO PQGCGQGYTPAQPAQGGQGYPASQPGQCGGHYLASQPGQCGQRHPASL QPRGCGQHYSATSLAQPGQCGGHYPASLQVGGQQTGLGQRQPGRGQOTROGQ QLPQGGQPGQGTROGLAQPGQCGQGYPTSPQSGQGGQPGQSGQPGQGGQ GVYSSLPQPGCLQGHYPASQPGQCGHGPGRQPGQCGQPGQCGQGYPT SPQPGQKGLQGGQGYPTSPQPGQCGQPGQCGQCPTSPQTGQAQPGQCGQ IGVQPGQCGQGYPTSLQSGQGGQGGHQLGQCGQSGQGGQGYDNPPIH VNTEQTASPKVAKVQAPATQLPIMCRMEGGDALSAQ"		

BASE COUNT      737 a    612 c    563 g    251 t

ORIGIN

Alignment Scores:

Pred. No.:	7.18e-07	Length:	2163
Score:	289.00	Matches:	85
Percent Similarity:	42.25%	Conservative:	24
Best Local Similarity:	32.95%	Mismatches:	47
Query Match:	25.17%	Indels:	102
DB:	8	Gaps:	13

US-10-053-410-4 (1-206) x AY245797 (1-2163)

QY	17	LeuValAlaValAlaValCysGlnGlyValGluArgGlnArgLeuArgAspLeuGln	36
Db	43	CTCGTGCTCTCACTGCTGCTGAAGGTGAGGCCTCTTAGGCCAA-----CTACAG	90
QY	37	CysTrpGlnGluValGlnGluSerProLeuAspAlaCysArgGlnValLeuAspArgGln	56
Db	91	TGTGAGCGCGAGCTCCAGGAGAGTCGCTTGAGGCATGCCACAGGTCTGTGGACCACAG	150
QY	57	LeuThrGlyGlyGlyGlyValGlyValGlyProPheargTrpGlyThrGlyLeuArg	76
Db	151	TTG-----GCCGCTGGCTGCATGGAGCACGGGGCTCCAG	186
QY	77	MetArgCysGlnGlnLeuGlnAspValSerArgGluCysArgCysAlaAlaLeuArg	96
Db	187	ATCGCATGCTGCCAGACAGCTCCGAGATGTTAGCGCTAAGTGCCGCTCTGTCGCGCTCAGC	246
QY	97	SerMetValargGlyTyrrGluGlu--AlaMetProProLeuGluIlysGlyTrpTrpPro	115
Db	247	CAAGTCGTAAGAACAATATGACAAACCGTGGTGGCGGCCAACAGGGCGGATCCTTCTACCT	306
QY	116	--TrpGlyArg-----	118
Db	307	GGCGAGACCACACACTGCAGCAACTCCAAACAAGATAATTATTTGGGAACATCTTCACAA	366

```

/evidence=experimental
/product="high molecular weight glutenin subunit y"
/protein_id="CAC40687.1"
/db_xref="GI:14329763"
/db_xref="GOA:O94IJ6"
/db_xref="SPTRMBL:O94IJ6"
/translations="MAKRLVLFATVITLVALTAAGEASRQLQCEIRELEQSSLEACR
QVVDQQLAGRLPSITGLQMRCCSLKRLVAVSVQVRYQETVTPVPPKGGSPY
GETTLPQLQQLVFWGTSQVGYPSVSPGQVPYPGQASPPQGPQGPQGPQGPQ
LGQCGQGYPTSLHSGQGGQGYPPSSLPQPGQOQIGQSGQGYPTSLQPGQGGQI
GQCGQGYPTSPQHPGQRQPGQGGQOQIGQSGQOQIGQSGQGGQGYPTSPQ
PLCGQGPQGWQGGQGGQGYPTSPQPGQGGQOQIPASQPGQGGQOQIPASQ
QPGQGGQYPASQPGQGGQOQIPASQPGQGGQOQIPASQPGQGGQOQIPASQ
QPGQGGQHYTASLPQGGQGGQHYTASLPQGGQGGQOQIGLQGRQPGQGGQOQ
QPGQGGQGGQGGQOQIGLQGRQPGQGGQGGQHYTASLPQGGQGGQOQIGLQGR
QLEQGGQGGQGGQOQIGLQGRQPGQGGQGGQHYTASLPQGGQGGQOQIGLQGR
GYSLSLQPGQGLQGHYPASLPQGGQGGQHYTASLPQGGQGGQOQIPGQGGQYPT
SPQPGQGGQKQGGQGGQGGQYPTSPQPGQGGQOQIPGQGGQOQIPGQGGQYPT
IGVQPGQGGQGGQYPTSPQPGQGGQOQIPGQGGQOQIPGQGGQOQIPGQGGQ
VNTQEQTASPKVAKVQOQATQPLPIMCRMEGGDALSASQ"
140..202
/sig_peptide
/mat_peptide
/gene="glu-1R"
/evidence=experimental
203..2299
/mat_peptide
/gene="glu-1R"
/product="HMW glutenin subunit y"
/evidence=experimental
BASE COUNT 794 a 668 c 583 g 285 t.
ORIGIN

Alignment Scores:
Pred. No.: 7,64e-07 Length: 2330
Score: 289.00 Matches: 85
Percent Similarity: 42.25% Conservative: 24
Best Local Similarity: 32.95% Mismatches: 47
Query Match: 25.17% Indels: 102
DB: 8 Gaps: 13

US-10-053-410-4 (1-206) x TAE314785 (1-2330)
QY 17 LeuValAlaValAlaValCysGInGlyValGluValGluArgLeuArgLeuGln 36
DB 182 CTGGTGGCTCTCACTGCTGCTGAAGGTGAGGCGCTCTAGGCAA-----CTACAG 229
QY 37 CysTrpGlnGluValGlnGluSerProLeuAspAlaCysArgGlnValLeuAspArgGln 56
DB 230 TGTGAGCGCAGCTCCAGGAGAGCTCGCTTGAGGCATGCCGACAGCTGTGGACCAACAG 289
QY 57 LeuThrGlyGlyGlyGlyGlyValGlyProPheArgTrpGlyThrGlyLeuArg 76
DB 290 TTG-----GCCGGTGGCTGCGCATGGAGCACGGGGCTCCAG 325
QY 77 MetArgCysCysGlnGlnLeuGlnAspValSerArgGluCysArgCysAlaAlaLeuArg 96
DB 326 ATGCGATGCTGCAGCAGCTCCGAGATGTTAGCGCTAAGTGCCTCTCGTCGCGCGTCAGC 385
QY 97 SerMetValArgGlyTyrGluGlu-----AlaMetProLeuGluGlyGlyTrpTrpPro 115
DB 386 CAAGTCGTATAGCAATATGAGCAACCGTGGTGGCGGCCCAAGGGCGGATCCTTCTACCT 445
QY 116 -----TrpGlyArg----- 118
DB 446 GCGGAGACCACACCACTGCAGCAATCCAAAGTAATATTTTGGGGAACATCTTCACAA 505
QY 119 -----GlnGlnPro----- 122
DB 506 ACAGTACAGGGTATTACCAAGCGTAAAGTTCTCTCAGCAGGGGCGCATATTATCCAGGC 565
QY 123 -----ProProGlnGlyGly-----Gly 128
DB 566 CAAGCTTCTCCAAACAGCCAGGACCAAGGCAACAGCCAGGCAATGGCAAGAACTGGGA 625
QY 129 GlyGlyGlnGlyTyr----- 134

```





Db 781 TACCATGTTAGCGGAGCAGCAGCGCGCGCTAAAGGTGGCAAGGCGCAGCAGCCC 840  
Qy 182 AlaAlaGlyLeuProMetMetCysArgLeuSerGluProGlnGluCysSerIlePheSer 201  
Db 841 GCGACACAGCTGTGCGCAATGTGCGGATG-----GAG 873  
Qy 202 GlyGlyAsp 204  
Db 874 GGGGGCGAC 882

Search completed: November 29, 2003, 18:16:13  
Job time : 3585 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: November 29, 2003, 17:10:31 ; Search time 2046 Seconds  
(without alignments)  
2447.078 Million cell updates/sec

Title: US-10-053-410-4  
Perfect score: 1148  
Sequence: 1 MAKIAAAAAALCFALVAV.....MMCRISPEOCESIFSGDQY 206

Scoring table: BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 22781392 seqs, 12152238056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:  
-MODEL=frame+ p2n.model -DRV=xlh  
-Q/cgn2\_1/USPTO.spool/US10053410/runat\_28112003\_140757\_27896/app\_query.fasta\_1.391  
-DB=EST -QFMT=fastap -SUFFIX=p2n.rst -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=bloum62 -TRANS=human40.cdi -LIST=45  
-DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US10053410 @CGN 1 1 2810 @runat\_28112003\_140757\_27896 -NCPU=6 -ICPU=3  
-NO\_MMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : EST.\*  
1: em\_estba.\*  
2: em\_esthum.\*  
3: em\_estin.\*  
4: em\_estnu.\*  
5: em\_estov.\*  
6: em\_estpl.\*  
7: em\_estro.\*  
8: em\_htc.\*  
9: gb\_est1.\*  
10: gb\_est2.\*  
11: gb\_htc.\*  
12: gb\_est3.\*  
13: gb\_est4.\*  
14: gb\_est5.\*  
15: em\_estfun.\*  
16: em\_estom.\*  
17: em\_gss\_hum.\*  
18: em\_gss\_inv.\*  
19: em\_gss\_pin.\*  
20: em\_gss\_vrt.\*  
21: em\_gss\_vrt.\*  
22: em\_gss\_fun.\*  
23: em\_gss\_mam.\*  
24: em\_gss\_mus.\*  
25: em\_gss\_pro.\*  
26: em\_gss\_rod.\*  
27: em\_gss\_phg.\*  
28: em\_gss\_vrl.\*  
29: gb\_gss1.\*

29: gb\_gss2.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	1148	100.0	900	11	AY105070	Zea mays
c	1148	100.0	935	29	CC349094	CG349094 OGUEN26TH
3	1122	97.7	1038	29	CC349105	CC349105 OGUEN26TV
c	857.5	74.7	674	14	CA401604	CA401604 EL01N0423
5	815.5	71.0	602	9	AI711821	AI711821 605064E02
6	772	67.2	586	10	BE917827	BE917827 OV1_7_F02
7	753.5	65.6	505	10	BE917909	BE917909 OV1_7_F02
8	746	65.0	539	9	AI745997	AI745997 605078E03
c	704	61.3	473	14	CD443922	CD443922 EL01N0432
c	625	54.4	560	9	AI712201	AI712201 605065F05
c	615	53.6	570	14	CA402183	CA402183 EL01N0432
12	547.5	47.7	425	10	EG048804	EG048804 OV1_23_CO
13	457.5	39.9	496	10	BG049097	BG049097 OV1_23_CO
14	432.5	37.7	683	9	AU164031	AU164031 AU164041
c	431	37.5	389	9	AI670620	AI670620 605035H03
16	431	37.5	712	9	AU163963	AU163963 AU163963
17	425.5	37.1	595	13	BQ172799	BQ172799 1008004_X
18	425.5	37.1	704	9	AU094496	AU094496 AU094496
19	422.5	36.8	697	9	AU094596	AU094596 AU094596
20	421.5	36.7	756	13	BQ06521	BQ06521 WHE3580_A
21	418	36.4	707	9	AU164041	AU164041 AU164041
22	415.5	36.2	687	9	AU163861	AU163861 AU163861
23	413.5	36.0	624	9	AU163975	AU163975 AU163975
24	413.5	36.0	687	9	AU163860	AU163860 AU163860
25	413	36.0	706	9	AU163839	AU163839 AU163839
26	411.5	35.8	677	9	AU094642	AU094642 AU094642
27	410.5	35.8	701	9	AU163968	AU163968 AU163968
28	409.5	35.7	710	9	AU164043	AU164043 AU164043
29	409.5	35.7	711	9	AU094571	AU094571 AU094571
30	409.5	35.7	711	9	AU094571	AU094571 AU094571
31	408.5	35.6	705	9	AU165812	AU165812 AU165812
32	407.5	35.5	527	9	AU076079	AU076079 AU076079
33	406.5	35.4	698	9	AU164035	AU164035 AU164035
34	406	35.4	707	9	AU164016	AU164016 AU164016
35	404.5	35.2	656	9	AU094570	AU094570 AU094570
36	404.5	35.2	688	9	AU163853	AU163853 AU163853
37	404.5	35.2	712	9	AU094563	AU094563 AU094563
38	402.5	35.1	650	9	AU091910	AU091910 AU091910
39	402	35.0	652	14	CA402267	CA402267 EL01N0434
40	401.5	35.0	675	9	AU094637	AU094637 AU094637
41	401	34.9	367	12	BM500544	BM500544 PAC000000
42	400.5	34.9	661	9	AU165797	AU165797 AU165797
43	399.5	34.8	643	9	AU164000	AU164000 AU164000
44	399.5	34.8	669	9	AU163999	AU163999 AU163999
45	399	34.8	708	9	AU163878	AU163878 AU163878

ALIGNMENTS

RESULT 1  
LOCUS AY105070 900 bp mRNA linear HTC 16-OCT-2002  
DEFINITION Zea mays PC0101751 mRNA sequence.  
ACCESSION AY105070  
VERSION AY105070.1 GI:21208148  
KEYWORDS HTC.  
SOURCE Zea mays  
ORGANISM Zea mays  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD  
clade; Panicoideae; Andropogoneae; Zea.  
REFERENCE 1 (bases 1 to 900)



**AUTHORS** Hayney, C.F., Dolan, M., Miao, G.H., Vogel, J.M., Whitsitt, M.S., Arthur, L.W., Hanafey, M., Morgante, M., and Tinsley, S.V.

**TITLE** Maize Mapping Project/DuPont Consensus Sequences for Design of Overgo Probes

**JOURNAL** Unpublished (2002)

**REFERENCE** 2 (bases 1 to 900)

**AUTHORS** Coe, E.H.

**TITLE** Direct Submission

**JOURNAL** Submitted (25-APR-2002) Maize Mapping Project, University of Missouri, Columbia, MO 65211, USA

**COMMENT** If you are interested in getting corresponding physical clones, these are publicly available from ZmDB and may be found by BLAST searching at MSU, maizegap.org; ZmDB, www.zmdb.iastate.edu; TIGR, www.tigr.org; or NCBI, www.ncbi.nlm.nih.gov. When the source of the maize cDNA sequences is either Virginia Walbot, Stanford or Pat Schnable, Iowa State, then clones may be requested from ZmDB: www.zmdb.iastate.edu.

**FEATURES** Location/Qualifiers

source

1..900

/organism="Zea mays"

/mol\_type="mRNA"

/db\_xref="maizeDB:635853"

/db\_xref="taxon:4577"

/clone\_lib="Maize Mapping Project/DuPont Consensus Library"

/note="this sequence is part of a project of EST assemblies resulting from the application of public contigs to seed DuPont contigs; this resource was assembled by DuPont as part of a collaboration for the overgo addressing of BACs in conjunction with the Maize Mapping Project"

BASE COUNT 191 a 244 c 331 g 134 t

ORIGIN

Alignment Scores:

Pred. No.: 4.92e-73 Length: 900

Score: 1148.00 Matches: 206

Percent Similarity: 100.00% Conservative: 0

Best Local Similarity: 100.00% Mismatches: 0

Query Match: 100.00% Indels: 0

DB: 11 Gaps: 0

US-10-053-410-4 (1-206) x AV105070 (1-900)

QY 1 MetAlaValLeuAlaAlaAlaAlaAlaLeuCysPheAlaAlaLeuValAlaVal 20

Db 68 ATGGCTAAGATCGCGCGCGCGCGCGCGCGCTGTGTTCGCGCCCTGTGTGGCGGTG 127

QY 21 AlaValCysGlnGlyGluValGluArgGlnArgLeuArgAspLeuGlnCysTrpGlnGlu 40

Db 128 GCCGTCTGCCAAGCGAGGTCGAGCGCGAGGCTCAGGACCTCAGTGTCTGGCAGGAG 187

QY 41 ValGlnGluSerProLeuAspAlaCysArgGlnValLeuAspArgGlnLeuThrGlyGly 60

Db 188 GTCCAGAGAGCCGCTCGAGCGGTGCCCGCAGGTCTTCGCGCGAGCTAACCGGGGC 247

QY 61 GlyGlyGlyGlyValGlyProPheArgTrpGlyLeuArgMetArgCysCys 80

Db 248 GGGCGGGCGCGCGGTGGCCGCTTCGCGCGCGCGCGCGCGCGCTCGGATCGCGTGC 307

QY 81 GlnGlnLeuGlnAspValSerArgGluCysArgCysAlaAlaLeuArgSerMetValArg 100

Db 308 CAGCAGCTCCAGAGCGTGAAGCGCGAGTGGCGCTGCGCGCGCCATCGGAGCATGGTCAGG 367

QY 101 GlyTyTrpGluGluAlaMetProProLeuGluGlyGlyTrpTrpProTrpGlyArgGlnGln 120

Db 368 GGCTACGAGGAGGAGCCATGCCCGCTCGAGAAAGGTGGTGGCCATGGGGCGGCGAGCAG 427

QY 121 GlnProProGlnGlyGlyGlyGlyGlyGlnGlyGlyTyTrpTyTrpProCysSerArg 140

Db 428 CAGCCCGCGCGCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 487

QY 141 ProGlyGluGlyTyTrpGlyTyGlyGlnGlyGlyGlnArgGlnMetTyTrpProCysArg 160

Db 488 CCAGGAGAGGGATACGGCTACGGTCAGGTGAGGTGCCAGCGGCAGATGTATCCACCGTGTCTGT 547

QY 161 ProGlyThrGlyGlyProArgGlnGlyArgValArgLeuThrLysAlaArgGlu 180

Db 548 CCGCGCACCCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 607

QY 181 TyrAlaAlaGlyLeuProMetMetCysArgLeuSerGluProGlnGluCysSerIlePhe 200

Db 608 TAGCCCGCGCGGTGCCGATGATGTCCGCGCTGTCCGAGCCCGCAGGATGCAGCATCTTC 667

QY 201 SerGlyGlyAspGlnTyR 206

Db 668 TCCGCGCGCGCAGCAGTAC 685

RESULT 2

CC349094/c

LOCUS

DEFINITION

OGUEN26TH ZM 0.7 1.5 KB Zea mays genomic clone ZMMBMA0422E04,

genomic survey sequence.

ACCESSION

CC349094

VERSION

CC349094.1 GI:30818501

KEYWORDS

GSS.

SOURCE

Zea mays

ORGANISM

Zea mays

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD clade; Panicoideae; Andropogoneae; Zea.

REFERENCE

1 (bases 1 to 935)

AUTHORS

Whitelaw, C.A., Quackenbush, J., Van Aken, S., Utterback, T., Resnick, A., Fraser, C.M., Budiman, M.A., Bedell, J.A., Rohlfing, T., Citek, R.W., Numberg, A., Robbins, D. and Lakey, N.

TITLE

Consortium for Maize Genomics

JOURNAL

Unpublished

COMMENT

Contact: Cathy Whitelaw

TIGR

9712 Medical Center Drive, Rockville, MD 20850, USA

Tel: 301-838-5843

Fax: 301-838-0208

Email: whitelaw@tigr.org

Seq primer: TR

Class: sheared ends.

FEATURES

Location/Qualifiers

source

1..935

/organism="Zea mays"

/mol\_type="genomic DNA"

/strain="B73"

/db\_xref="taxon:4577"

/clone="ZMMBMA0422E04"

/clone\_lib="ZM\_0.7 1.5 KB"

/note="Vector: pBCSK-; Site 1: HincII; 0.7-1.5 kb methylation filtered genomic DNA library"

BASE COUNT 147 a 339 c 259 g 190 t

ORIGIN

Alignment Scores:

Pred. No.: 5.12e-73 Length: 935

Score: 1148.00 Matches: 206

Percent Similarity: 100.00% Conservative: 0

Best Local Similarity: 100.00% Mismatches: 0

Query Match: 100.00% Indels: 0

DB: 29 Gaps: 0

US-10-053-410-4 (1-206) x CC349094 (1-935)

QY 1 MetAlaValLeuAlaAlaAlaAlaAlaLeuCysPheAlaAlaLeuValAlaVal 20

Db 834 ATGGCTAAGATCGCGCGCGCGCGCGCGCGCTGTGTTCGCGCCCTGTGTGGCGGTG 775

QY 21 AlaValCysGlnGlyGluValGluArgGlnArgLeuArgAspLeuGlnCysTrpGlnGlu 40

Db 774 GCCGTCTGCCAAGCGAGGTCGAGCGCGCAGAGGCTCAGGGA CCTGCAGTGTCTGGCAGGAG 715

```

QY 41 ValGlnGluSerProLeuAspAlaCysArgGlnValLeuAspArgGlnLeuThrGlyGly 60
Db 714 GTCCAGAGAGCCCGCTCGACGCGTCCCGCAGGTCTCGACCGGAGCTAACCGCGCG 655

QY 61 GlyGlyGlyGlyValGlyProPheArgTgGlyThrGlyLeuArgMetArgCysCys 80
Db 654 GCGCGCGCGCGCGCGTTCGCCGTTCCGGTGGGGCACCAGGCTCCGGATGCGGTCTGC 595

QY 81 GlnGlnLeuGlnAspValSerArgGluCysArgCysAlaAlaIleArgSerMetValArg 100
Db 594 CAGCAGCTCCAGGACGTGAGCGCGAGTGCCGCTGCGCGCCATCCGAGCATGTGCAGG 535

QY 101 GlyTyGlnGluAlaMetProProLeuGluLysGlyTyrTrpProTrpGlyArgGlnGln 120
Db 534 GGTACAGAGAGCCATGCGCGCTCGGAGAAAGGCTGGTGGCATATGACCCCTGCAGCGAG 475

QY 121 GlnProProGlnGlnGlyGlyGlyGlnGlnGlyGlyTyrTyrProCysSerArg 140
Db 474 CAGCCCGCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 415

QY 141 ProGlyGluGlyTyrGlyTyrGlyGlnGlyGlyGlnArgGlnMetTyrProCysArg 160
Db 414 CCAGGAGAGGATACCGCTACGGTCAAGGTGGCCAGCGGAGATGTATCCACCGTCTGT 355

QY 161 ProGlyThrGlyGlyProArgIleGlyArgValArgLeuThrLysAlaArgGlu 180
Db 354 CCGCGCACCCAGCGCGCGCGGAGGATCGCGCGCGGTGAGGCTTACGAAGGCGCGGAG 295

QY 181 TyrAlaAlaGlyLeuProMetMetCysArgLeuSerGluProGlnGluCysSerIlePhe 200
Db 294 TAGCGCGCGGCGTTCGCGATGTGTCGCGCTGTGCGAGCCCGGAGGATGCAGCATCTTC 235

QY 201 SerGlyGlyAspGlnTyr 206
Db 234 TCCGCGCGGCGACCACTATC 217

RESULT 3
CC349105 1038 bp DNA linear GSS 16-MAY-2003
LOCUS OGUEN26TV_ZM_0.7_1.5_KB_Zea_mays_genomic_clone_ZM58Wa0422E04,
DEFINITION genomic survey sequence.
ACCESSION CC349105
VERSION CC349105.1 GI:30818512
KEYWORDS GSS.
SOURCE Zea mays
ORGANISM Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
clade; Panicoideae; Andropogoneae; Zea.
Whitehead, C.A., Quackenbush, J., Van Aken, S., Utterback, T., Resnick
, A., Fraser, C.M., Budiman, M.A., Bedell, J.A., Rohlfing, T., Citek
, R.W., Nunberg, A., Robbins, D. and Lakey, N.
Consortium for Maize Genomics
Unpublished
Contact: Cathy Whitelaw
TIGR
9712 Medical Center Drive, Rockville, MD 20850, USA
Tel: 301-838-5943
Fax: 301-838-0208
Email: whitelaw@tigr.org
Seq primer: TP
Class: sheared ends.
FEATURES
    source
        1..1038
            /organism="Zea mays"
            /mol_type="genomic DNA"
            /strain="B73"
            /db_xref="taxon:4577"
            /clone_lib="ZM_0.7_1.5_KB"
            /clone="ZM58Wa0422E04"
            /notes="Vector: pBCSK-; Site 1: HincII; 0.7-1.5 kb
            methylation filtered genomic DNA library"

```

```

BASE COUNT 215 a 316 c 336 g 171 t
ORIGIN
Alignment Scores:
Pred. No.: 4,18e-71 Length: 1038
Score: 1122.00 Matches: 206
Percent Similarity: 99.04% Conservative: 0
Best Local Similarity: 99.04% Mismatches: 0
Query Match: 97.74% Indels: 2
DB: 29 Gaps: 0
US-10-053-410-4 (1-206) x CC349105 (1-1038)
QY 1 MetAlaLysIleAlaAlaAlaAlaAlaAlaLeuCysPheAlaAlaVal 20
Db 393 ATGGCTAAGATCGCCGCGCGCGCGCGCGCTGTGCTTCGCGGCCCTGGTGGCGGTG 452
QY 21 AlaValCysGlnGlyValGluValGluArgGlnArgLeuArgAspLeuGlnCysTrpGlnGlu 40
Db 453 GCCGTCTGCAAGGCGAGGTCCAGGCGGAGAGGCTCAGGAGCTCAGTGTGCTGCGCAGGAG 512
QY 41 ValGlnGluSerProLeuAspAlaCysArgGlnValLeuAspArgGlnLeuThrGlyGly 60
Db 513 GTCCAGGAGAGCCCGCTCGACGCGTCCGCCAGGTCCTCGACCGGAGCTAACCGCGCGC 572
QY 61 GlyGlyGlyGlyValGlyValGlyProPheArgTgGlyThrGlyLeuArgMetArgCysCys 80
Db 573 GCGCGCGCGCGCGGTGGCCGCTTCGGTGGGGCACCAGGCTCCGGATGCGGTCTGCTGC 632
QY 81 GlnGlnLeuGlnAspValSerArgGluCysArgCysAlaAlaIleArgSerMetValArg 100
Db 633 CAGCAGCTCCAGGACGTGAGCGCGAGTGCCTGCGCGCCATCCGAGCATGTGTCTCAGG 692
QY 101 GlyTyGlnGluAlaMetProProLeuGluLysGlyTyrTrpProTrpGlyArgGlnGln 120
Db 693 GGCTACGAGAGGCGCATGCCCGCTGGAGAAAGGCTGGTGGCGCATGGGCGCGGAGCAG 752
QY 121 GlnProProGlnGlnGlyGlyGlyGlnGlnGlyTyrTyrProCysSerArg 140
Db 753 CAGCGCGCGCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 812
QY 141 ProGlyGluGlyTyrGlyTyrGlyGlnGlyGlnArgGlnMetTyrProCysArg 160
Db 813 CCAGGAGAGGATACCGCTACGGTCAAGGTGGCGCGGAGATGTATCCACCGTCTGT 872
QY 161 ProGlyThrGlyGlyProArgIleGlyArgValArgLeuThrLysAlaArg-Gl 180
Db 873 CCGCGCACCCAGCGCGCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 932
QY 180 uTyrAlaAlaGlyLeuProMetMetCysArgLeuSerGluProGlnGluCysSerIlePh 200
Db 933 GTACGCGCGCGGTTGCCGATGATGTGCCGCTGTCCGAGCCCCCAGGAGTGCAGCATCTT 992
QY 200 eSerGly-GlyAspGlnTyr 206
Db 993 CTCGCGCGCGCGACCACTATC 1012

RESULT 4
CA401604/c
LOCUS ELOIN0423A08.g Endosperm_4 Zea mays cDNA, mRNA linear EST 07-NOV-2002
DEFINITION CA401604
ACCESSION CA401604
VERSION CA401604.1 GI:24766450
KEYWORDS EST.
SOURCE Zea mays
ORGANISM Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
clade; Panicoideae; Andropogoneae; Zea.
REFERENCE 1 (bases 1 to 674)
AUTHORS Lai, J., Dey, N., Kim, C.S., Becraft, P., Larkins, B., Linton, E. and
Messing, J.
TITLE Sequencing of the maize endosperm ESTs

```



```

Db      189 CCCAGAGTGCAGCGTCTTCTCCGGCGGACCAAGTAC 151
RESULT 6
BE917827
LOCUS   BE917827          586 bp      mRNA      linear      EST 29-SEP-2000
DEFINITION OVI_7_F02.g1_A002 Ovary 1 (OVI) sorghum bicolor cDNA, mRNA
sequence.
ACCESSION BE917827
VERSION   BE917827.1 GI:10420322
KEYWORDS  Sorghum bicolor (sorghum)
SOURCE    Sorghum bicolor
ORGANISM  Sorghum bicolor (sorghum)

REFERENCE
AUTHORS   Cordonnier-Pratt,M.-M., Gingle,A., Marsala,C., Sudman,M. and Pratt
           ,L.H.
TITLE     An EST database from Sorghum: ovaries of varying immature stages
JOURNAL   Unpublished
COMMENT   Contact: Cordonnier-Pratt MM
           Laboratory for Genomics and Bioinformatics
           The University of Georgia, Department of Plant Biology
           Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
           Tel: 706 542 1860
           Fax: 706 583 0210
           Email: mmpratt@uga.edu
           Sequences have been trimmed to exclude PolyA, vector and regions
           below Phred quality 16. The threshold for highest quality sequence
           is 20.
           Seq primer: PolyTWix
           High quality sequence start: 3
           High quality sequence stop: 584
           POLYA=No.
           Location/Qualifiers
             source          1..586
                           /organism="Sorghum bicolor"
                           /mol_type="mRNA"
                           /db_xref="taxon:4558"
                           /clone_lib="Ovary 1 (OVI)"
                           /notes="Organ: Mix of ovaries of varying immature stages
                           from 8-week-old plants; Vector: pBluescript II from Lambda
                           Zap II; Site 1: XhoI; Site 2: EcoRI; The library was made
                           from poly-A RNA in the cloning vector lambda Zap II.
                           Clones to be sequenced were prepared by mass excision."

BASE COUNT  115 a  179 c  212 g   80 t

FEATURES
source
  1..586
    /organism="Sorghum bicolor"
    /mol_type="mRNA"
    /db_xref="taxon:4558"
    /clone_lib="Ovary 1 (OVI)"
    /notes="Organ: Mix of ovaries of varying immature stages
    from 8-week-old plants; Vector: pBluescript II from Lambda
    Zap II; Site 1: XhoI; Site 2: EcoRI; The library was made
    from poly-A RNA in the cloning vector lambda Zap II.
    Clones to be sequenced were prepared by mass excision."

Alignment Scores:
Pred. No.:      3.07e-46      Length:      586
Score:          772.00      Matches:      144
Percent Similarity: 84.32%      Conservative: 12
Best Local Similarity: 77.84%      Mismatches: 13
Query Match:    67.25%      Indels:      16
DB:             10          Gaps:         4

US-10-053-410-4 (1-206) x BE917827 (1-586)

Qy      22 ValCysGlnGluValGluArgGlnArgLeuArgAspLeuGlnCysTrpGlnGluVal 41
Db      2  GTCCGCCAAGCGCTGGTCGAGCGCGCAGAGCTCAAGACCTCGAGTCTGCGCAAGAGGTC 61
Qy      42 GlnGluSerProLeuAspAlaCysArgGlnValLeuAspArgGlnLeuThrGlyGly 61
Db      62 CAGAGAACCCTCGCGCGTCCGGCAGGTCTCGACCGCAGTCTACCC---GGCGGC 118
Qy      62 GlyGlyGlyValGlyProPheArgTrpGlyThrGlyLeuArgMetArgCysGln 81
Db      119 ATCGGTACGCATCGCCCGCTCCGGTGGGGCACCAGGCTCCGGATGGGTCTCCAG 178
Qy      82 GlnLeuGlnAspValSerArgGluCysArgCysAlaAlaIleArgSerMetValArgGly 101
Db      179 CAGCTCCAGACGTCAGCCGCGAGTGCCTGCGCGCCATCCGAGCATGGTGGTGGGC 238

```

---

```

Qy      102 TYrGluGluAlaMetProProLeuGluLysGlyTrpTrpProTrpGlyArgGlnGlnGln 121
Db      239 TACGAGGAGACCATGCCCGCGCTGGAGAAAGCTGG-----TGGGGGCGAGCCGCGAG 292
Qy      122 ProProProGlnGlnGlyGlyGlyGlnGlyGlyTrpTrpTyProCysSerArgPro 141
Db      293 CCG-----GGCTACGACTACCCCTGCAGCCAGCA 322
Qy      142 GlyGluGlyTrpGlyTrpGlyGlnGlnArgGlnMetTyProProCysArgPro 161
Db      323 GGAGAGGGATACGGTACGTGAGAGTGCAGCAGCAGATGTACCCACCGTGTCTCTCC 382
Qy      162 GlyThrThrGlyGlyProArgIleGlyArgValArgLeuThrLysAlaArgGlyTrp 181
Db      383 GGCACC-----GGGCAAAAATCGCCCGTGAAGCTTACAAAGGCCAGGCAGTAC 433
Qy      182 AlaAlaGlyLeuProMetMetCysArgLeuSerGluProGlnGluCysSerIlePheSer 201
Db      434 GCCGCGAGGATCCGATGATGTCCGCGTGTCCGAGCCCCCAGAGTGCAGCGTCTTCTCC 493
Qy      202 GlyGlyAspGlnTyr 206
Db      494 GCGCGCGACCAAGTAC 508

RESULT 7
BE917909
LOCUS   BE917909          605 bp      mRNA      linear      EST 29-SEP-2000
DEFINITION OVI_7_F02.bl_A002 Ovary 1 (OVI) sorghum bicolor cDNA, mRNA
sequence.
ACCESSION BE917909
VERSION   BE917909.1 GI:10420483
KEYWORDS  Sorghum bicolor (sorghum)
SOURCE    Sorghum bicolor
ORGANISM  Sorghum bicolor (sorghum)

REFERENCE
AUTHORS   Cordonnier-Pratt,M.-M., Gingle,A., Marsala,C., Sudman,M. and Pratt
           ,L.H.
TITLE     An EST database from Sorghum: ovaries of varying immature stages
JOURNAL   Unpublished
COMMENT   Contact: Cordonnier-Pratt MM
           Laboratory for Genomics and Bioinformatics
           The University of Georgia, Department of Plant Biology
           Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
           Tel: 706 542 1860
           Fax: 706 583 0210
           Email: mmpratt@uga.edu
           Sequences have been trimmed to exclude PolyA, vector and regions
           below Phred quality 16. The threshold for highest quality sequence
           is 20.
           Seq primer: JEN REV
           High quality sequence stop: 523
           POLYA=No.
           Location/Qualifiers
             source          1..605
                           /organism="Sorghum bicolor"
                           /mol_type="mRNA"
                           /db_xref="taxon:4558"
                           /clone_lib="Ovary 1 (OVI)"
                           /notes="Organ: Mix of ovaries of varying immature stages
                           from 8-week-old plants; Vector: pBluescript II from Lambda
                           Zap II; Site 1: XhoI; Site 2: EcoRI; The library was made
                           from poly-A RNA in the cloning vector lambda Zap II.
                           Clones to be sequenced were prepared by mass excision."

BASE COUNT  122 a  184 c  217 g   82 t

FEATURES
source
  1..605
    /organism="Sorghum bicolor"
    /mol_type="mRNA"
    /db_xref="taxon:4558"
    /clone_lib="Ovary 1 (OVI)"
    /notes="Organ: Mix of ovaries of varying immature stages
    from 8-week-old plants; Vector: pBluescript II from Lambda
    Zap II; Site 1: XhoI; Site 2: EcoRI; The library was made
    from poly-A RNA in the cloning vector lambda Zap II.
    Clones to be sequenced were prepared by mass excision."

Alignment Scores:
Pred. No.:      6.74e-45      Length:      605
Score:          753.50      Matches:

```

```

Percent Similarity: 83.16%
Best Local Similarity: 77.37%
Query Match: 65.64%
DB: 10
Conservative: 11
Mismatch: 15
Indels: 17
Gaps: 5

US-10-053-410-4 (1-206) x BE917909 (1-605)
QY 1 MetAlaLysIleAlaAla---AlaAlaAlaAlaLeuCysPheAlaAlaLeuValAla 19
DB 82 ATGGCTAAGATCGTCCGCGCCGCGCCGCGCTGCTCGCGGCGCTGCTCGCGGCGCTGGGCC 141
QY 20 ValAlaValCysGlnGlyValGluValGluArgGlnArgLeuArgPheLeuGlnCysTrpGln 39
DB 142 GTGGCGCTCGGCCAAGCGTGGTCCGAGCGCGCAGAGGCTCAAGAGCCTCGCAGTGTGGCAA 201
QY 40 GluValGlnGlnSerProLeuAspAlaCysArgGlnValLeuAspArgGlnLeuThrGly 59
DB 202 GAGGTCGAGAGAAGACCGCTCGCGCGTGCAGGCTGCTCGACCGGCGAGCTCACC--- 258
QY 60 GlyGlyGlyGlyGlyValGlyProPheArgTrpGlyThrGlyLeuArgMetArgCys 79
DB 259 GCGCGGATGCGCTACGCGATCGGCCCTTCCGTGGGGCACCGGGCTCCGGATCGCGTGC 318
QY 80 CysGlnGlnLeuGlnAspValSerArgGluCysArgCysAlaAlaIleArgSerMetVal 99
DB 319 TGCAGCAGCTCCAGGACGTGAGCGCGAGTGCCTGCGCGCCATCCGGAGCATGGTC 378
QY 100 ArgGlyTrpGluGluAlaMetProProLeuGluLysGlyTrpTrpProTrpGlyArgGln 119
DB 379 AGGGGCTACGAGGAGACCATCGCGCGCTGCGGAGAAAGGCTGG-----TGGGGGCGCAG 432
QY 120 GlnGlnProProGlnGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGly 139
DB 433 CCGCAGCGG-----GGCTACGACTACCCCTCGAGC 462
QY 140 ArgProGlyGluGlyTrpGlyTrpGlyGlnGlyGlyGlnArgGlnMetTrpProCys 159
DB 463 CAGCGAGGAGAGGAGATACGCTACGCTGAGAGTGGCCAGCAGCAGATGATCCACCGTGT 522
QY 160 ArgProGlyThrThrGlyGlyGlyProArgIleGlyArgValArgLeuThrLysAlaArg 179
DB 523 CGTCCCGGCACC-----GGGCAAAAATCCCGCGCTGAAGCTTACAAAGGCCAGG 573
QY 180 GluTrpAlaAlaGlyLeuProMetMetCys 189
DB 574 CAGTACCGCCAGGATGCCGATGATGTGC 603

RESULT 8
AI745997/c
LOCUS AI745997 539 bp mRNA linear EST 02-FEB-2000
DEFINITION 605078E03.x1 605 - Endosperm cDNA library from Schmidt lab Zea mays
cDNA, mRNA sequence.
ACCESSION AI745997
VERSION AI745997.1 GI:5124261
KEYWORDS EST.
SOURCE Zea mays
ORGANISM Zea mays

REFERENCE
AUTHORS Walbot V.
TITLE Maize ESTs from various cDNA libraries sequenced at Stanford
JOURNAL University
COMMENT Unpublished
Department of Biological Sciences
Stanford University
855 California Ave, Palo Alto, CA 94304, USA
Tel: 650 723 2227
Fax: 650 725 8221
Email: walbot@stanford.edu
Plate: 605078 row: E column: 03.

FEATURES
source
Location/Qualifiers
1..539
/organism="Zea mays"
/mol_type="mRNA"
/db_xref="taxon:4577"
/tissue_type="nucellar, embryo, and endosperm"
/dev_stage="10-14 days post-pollination"
/lab_host="PH5(alpha)"
/clone_lib="605 - Endosperm cDNA library from Schmidt lab"
/note="Organ: Kernel; Vector: pAD-GAL4-2; Site: EcoRI; Site 2: XhoI; Kernel endosperm cDNA library from Schmidt lab"

BASE COUNT 79 a 204 c 151 g 105 t
ORIGIN
Alignment Scores: 2,07e-44 Length: 539
Pred. No.: 746.00 Matches: 131
Score: 97.78% Conservative: 1
Percent Similarity: 97.04% Mismatches: 1
Best Local Similarity: 64.98% Indels: 2
Query Match: 9 Gaps: 1
DB:

US-10-053-410-4 (1-206) x AI745997 (1-539)
QY 74 GlyLeuArgMetArgCysCysGlnGlnLeuGlnAspValSerArgGluCysArgCysAla 93
DB 538 GGGCTCCGATCGGTGCTGCCAGCAGCTCCAGGACGTGAGCGCGCGAGTGCCTCGGCC 479
QY 94 AlaIleArgSerMetValArgGlyTrpGluGluAlaMetProProLeuGluLysGlyTrp 113
DB 478 GCCATCCGAGCATGTGTGAGGGCTACGAGGAGGCCATCCCGCGCGCGAGAAAGCTGG 419
QY 114 TrpProTrpGlyArgGlnGlnProProGln-----GlyGlyGlyGlyGln 131
DB 418 TGGCCATGGGGCGCGCAGCAGCAGCGCGCGCGAGGAGGAGGAGGAGGAGGAGGAGCAG 359
QY 132 GlyGlyTrpTrpTrpProCysSerArgProGlyGlyGlyTrpGlyTrpGlyGlnGlyGly 151
DB 358 GGGGGCTACTACTACCCCTGCGAGCGCGCGCAGGAGGAGGATCGGCTACGGTACGGTGGC 299
QY 152 GlnArgGlnMetTrpProCysArgProGlyTrpThrGlyGlyProArgIleGly 171
DB 298 CAGCGGCAGATGTATCCACCTGTGTCGCGGACCAACCGCGCGCGGCGGCGGCGGCGG 239
QY 172 ArgValArgLeuThrLysAlaArgGluTrpAlaAlaGlyLeuProMetMetCysArgLeu 191
DB 238 CGCGTACGGCTTACGAAGCGCGGAGTACGCGCGCGGTTGCGGATGATGTGCGCGCTG 179
QY 192 SerGluProGlnGluCysSerIlePheSerGlyGlyAspGlnTyr 206
DB 178 TCGGAGCCCGAGGATGCGCTCTTCTCCGCGCGCGCAGGTAC 134

RESULT 9
CD443922
LOCUS CD443922 473 bp mRNA linear EST 03-JUN-2003
DEFINITION EL01N0432G05.b Endosperm_4 Zea mays cDNA, mRNA sequence.
ACCESSION CD443922
VERSION CD443922.1 GI:31359565
KEYWORDS EST.
SOURCE Zea mays
ORGANISM Zea mays

REFERENCE
AUTHORS Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
clade; Panicoideae; Andropogoneae; Zea.
1 (bases 1 to 473)
AUTHORS Lai, J., Dey, N., Kim, C.S., Becraft, P., Larkins, B., Linton, E. and
Messing, J.
TITLE Sequencing of the maize endosperm ESTs
JOURNAL Unpublished
COMMENT Contact: Lai, Jinseng
Dr. Joachim Messing's lab

```

Waksman Institute, Rutgers University  
130 Frelinghuysen Rd., Piscataway, NJ 08854, USA  
Tel: 732-445-3801  
Fax: 732-445-5735  
Email: jlai@waksman.rutgers.edu  
Seq primer: T3.

## FEATURES

source

Location/Qualifiers  
1. 473  
/organism="Zea mays"  
/mol\_type="mRNA"  
/cultivar="W22"  
/db\_xref="taxon:4577"  
/tissue\_type="Endosperm of 7-23DAP"  
/clone\_lib="Endosperm\_4"  
/notes="Vector: pBluescript SK-; Site\_1: EcoRI; Site\_2: XhoI"

BASE COUNT 76 a 138 c 201 g 58 t  
ORIGIN

## Alignment Scores:

Pred. No.: 1.87e-41 Length: 473  
Score: 704.00 Matches: 131  
Percent Similarity: 94.96% Conservative: 1  
Best Local Similarity: 94.24% Mismatches: 3  
Query Match: 61.32% Indels: 4  
DB: 14 Gaps: 2

US-10-053-410-4 (1-206) x CD443922 (1-473)

Qy 1 MetAlaLysIle---AlaAlaAlaAlaAlaAlaLeuCysPheAlaAlaLeuValAla 19  
Db 66 ATGGCTAAGATCGCGCGCGCGCGCGCGCGCTGTGCTTCGGCGGCTCGTGGCC 125  
Qy 20 ValAlaValCysGlnGlyGluValGluArgGlnArgLeuArgAspLeuGlnCysTrpGln 39  
Db 126 GTGGCCCTCTGCCAAGCGAGGCTGACGCGCAGAGGCTCAGGGCGCGGAGTCTGCCAG 185  
Qy 40 GluValGlnGluSerProLeuAspAlaCysArgGlnValLeuAspArgGlnLeuThrGly 59  
Db 186 GAGGGCCAGGAGAGCCGCTCGACGCTGCGCGCAGGCTCTCGACGGCAGCTAAC--- 242  
Qy 60 GlyGlyGlyGlyGlyValGlyProPheArgTrpGlyThrGlyLeuArgMetArgCys 79  
Db 243 -----GGCGCGCGCGCTCGCGCGCTTCGGTGGGGCACCGGGCTCGCGATCGCGTGC 296  
Qy 80 CysGlnGlnLeuGlnAspValSerArgGluCysArgCysAlaAlaAlaAlaAlaAlaVal 99  
Db 297 TGCACAGCAGCTCCAGGACGTGAGCGCGAGTGCCTGCGCGCCATCCGGAGCATGGTC 356  
Qy 100 ArgGlyTyrgluGluAlaMetProLeuGluLysGlyTrpTrpProTrpGlyArgGln 119  
Db 357 AGGGCTACGAGAGGCGCATCGCGCTCTGAGANAGCTGTGGCCATGGGGCGGCGAG 416  
Qy 120 GlnGlnProProGlnGlyGlyGlyGlyGlyGlnGlyGlyGlyTyrtyrtyrProCys 138  
Db 417 CAGCAGCGCGCGCGCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 473

## RESULT 10

AT172201/c

LOCUS

605065F05.x1 605 - Endosperm cDNA library from Schmidt lab Zea mays  
cDNA, mRNA sequence.

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD  
clade; Panicoideae; Andropogoneae; Zea.

REFERENCE

AUTHORS

TITLE

Maize ESTs from various cDNA libraries sequenced at Stanford

## JOURNAL

COMMENT

University  
Unpublished  
Contact: Walbot V  
Department of Biological Sciences  
Stanford University  
855 California Ave, Palo Alto, CA 94304, USA  
Tel: 650 723 2227  
Fax: 650 725 8221  
Email: walbot@stanford.edu  
Plate: 605065 row: F column: 05.

## FEATURES

source

Location/Qualifiers  
1. 560  
/organism="Zea mays"  
/mol\_type="mRNA"  
/cultivar="Ohio43"  
/db\_xref="taxon:4577"  
/tissue\_type="nucellar, embryo, and endosperm"  
/dev\_stage="10-14 days post-pollination"  
/lab\_host="DH5(alpha)"  
/clone\_lib="605 - Endosperm cDNA library from Schmidt lab"  
/note="Organ: Kernel; Vector: pAD-GAL4-2; Site 1: EcoRI;  
Site 2: XhoI; Kernel endosperm cDNA library from Schmidt  
lab"

BASE COUNT 89 a 197 c 159 g 114 t 1 others  
ORIGIN

## Alignment Scores:

Pred. No.: 1.04e-35 Length: 560  
Score: 625.00 Matches: 113  
Percent Similarity: 89.15% Conservative: 2  
Best Local Similarity: 87.60% Mismatches: 12  
Query Match: 54.44% Indels: 3  
DB: 9 Gaps: 1

US-10-053-410-4 (1-206) x AT172201 (1-560)

Qy 80 CysGlnGlnLeuGlnAspValSerArgGluCysArgCysAlaAlaAlaAlaAlaAlaVal 99  
Db 560 TGCCACCACTCCAGGACGCTACCCCTC-GAGTGCCGNTGCCCTCCCTCCATCCGAGCATGGTC 502  
Qy 100 ArgGlyTyrgluGluAlaMetProLeuGluLysGlyTrpTrpProTrpGlyArgGln 119  
Db 501 AGGGGTTCAGAGGACCATGCCCCCGCGAGAAAGGCTGTGTGGCATGGGGCGGCGAG 442  
Qy 120 GlnGlnProProGln-----GlyGlyGlyGlyGlyGlnGlyGlyTyrtyrtyrPro 137  
Db 441 CAGACGCCCGCGCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 382  
Qy 138 CysSerArgProGlyGlyTyrglyTyrglyGlyGlnGlyGlnArgGlnMetTyrtyrPro 157  
Db 381 TGCAGCGCGCCAGGAGGAGGATACGGCTACGGTCAGGGTGGCCACCGGAGATGTATCCC 322  
Qy 158 ProCysArgProGlyThrThrGlyGlyGlyProArgIleGlyArgValArgLeuThrLys 177  
Db 321 CCTGTGCTCCGCGCCCCCGCGCGCGCAAGGATCGGCGCTGAGGCTTACGAAG 262  
Qy 178 AlaArgGluTyrgluAlaAlaGlyLeuProMetMetCysArgLeuSerGluProGlnGluCys 197  
Db 261 GCCCGGAGTATCCCCCGGGGTTCGCGATGATGTGCCGCTGTCCGAGCCCGGAGGTGC 202  
Qy 198 SerIlePheSerGlyGlyAspGlnTyrtyr 206  
Db 201 AGCGTCTTCTCCGCGCGCGCCAGTAC 175

## RESULT 11

CA402183/c

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

CA402183 570 bp mRNA linear EST 07-NOV-2002  
EL01N0432G05.g Endosperm\_4 Zea mays cDNA, mRNA sequence.

CA402183

GI:24767038

EST.

Zea mays

Zea mays





RESULT 13  
 BG049097  
 LOCUS  
 DEFINITION OVI\_23\_C05\_g1\_A002 Ovary 1 (OVI), Sorghum bicolor cDNA, mRNA  
 EST 25-JAN-2001  
 ACCESSION BG049097  
 VERSION BG049097.1 GI:12500510  
 KEYWORDS EST.  
 SOURCE Sorghum bicolor (sorghum)  
 ORGANISM Sorghum bicolor  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD  
 clade; Panicoideae; Andropogoneae; Sorghum.  
 1 (bases 1 to 496)  
 Cordonnier-Pratt,M.-M., Gingle,A., Marsala,C., Sudman,M. and Pratt  
 L.H.  
 TITLE An EST database from Sorghum: ovaries of varying immature stages  
 JOURNAL Unpublished  
 COMMENT Contact: Cordonnier-Pratt MM  
 Laboratory for Genomics and Bioinformatics  
 The University of Georgia, Department of Plant Biology  
 Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA  
 Tel: 706 542 1860  
 Fax: 706 583 0210  
 Email: mmpratt@uga.edu  
 Sequences have been trimmed to exclude PolyA, vector and regions  
 below Phred quality 16. The threshold for highest quality sequence  
 is 20.  
 Seq primer: PolyTwix  
 High quality sequence start: 43  
 High quality sequence stop: 482  
 POLYA=No.

## FEATURES

Location/Qualifiers  
 1..496

/organism="Sorghum bicolor"  
 /mol\_type="mRNA"

/db\_xref="taxon:4558"  
 /clone\_lib="Ovary 1 (OVI)"

/note="Organ: Mix of ovaries of varying immature stages  
 from 8-week-old plants; Vector: pBluescript II from Lambda  
 Zap II; Site\_1: XhoI; Site\_2: EcoRI; The library was made  
 from poly-A RNA in the cloning vector lambda ZAP II.  
 Clones to be sequenced were prepared by mass excision."

BASE COUNT 125 a 124 c 160 g 87 t  
 ORIGIN

## Alignment Scores:

Pred. No.: 9,63e-24 Length: 496  
 Score: 457.50 Matches: 84  
 Percent Similarity: 80.87% Conservative: 9  
 Best Local Similarity: 73.04% Mismatches: 7  
 Query Match: 39.85% Indels: 15  
 DB: 10 Gaps: 3

US-10-053-410-4 (1-206) x BG049097 (1-496)

QY 92 CysAlaAlaIleArgSerMetValArgGlyTyrGluGluAlaMetProLeuGluTys 111  
 Db 2 TCGCGCGCCATCCGGAGCATGGTCAGGGGCTAGAGGAGACCATCCGCGCTGGAGAAA 61  
 QY 112 GlyTyrTrpProTyrArgGlnGlnProProGlnGlyGlyGlyGln 131  
 Db 62 GACTGG-----TGGGGGAGCAGCGGAGCCG----- 88  
 QY 132 GlyGlyTyrTyrProCysSerArgProGlyGluGlyTyrGlyTyrGlyGlnGlyGly 151  
 Db 89 ---GGCTACGACTACCCCTGCAGCCAGGAGGAGGATACGGCTACGGTGAGAGTGGC 145  
 QY 152 GlnArgGlnMetTyrProProCysArgProGlyThrThrGlyGlyProArgIleGly 171  
 Db 146 CAGCAGCAGATGTACCCACCGTGTCTCCGCGCACC-----GGGCAAAAAATCGCC 196  
 QY 172 ArgValArgLeuThrIysAlaArgGluTyrAlaAlaGlyLeuProMetMetCysArgLeu 191

Db 197 CGCGTGAAGCTTACAAAGCCAGGAGTACGCGCAGGATCCGATGATGTCGCGCTG 256  
 QY 192 SerGluProGlnGluCysSerIlePheSerGlyGlyAspGlnTyr 206  
 Db 257 TCGGAGCCCCAGGAGTGCAGCGTCTTCCGCGCGGACCACTAC 301  
 RESULT 14  
 AUI64031  
 LOCUS  
 DEFINITION AUI64031 Rice panicle at ripening stage Oryza sativa (japonica  
 cultivar-group) cDNA clone E11772, mRNA sequence.  
 ACCESSION AUI64031  
 VERSION AUI64031.1 GI:11171929  
 KEYWORDS EST.  
 SOURCE Oryza sativa (japonica cultivar-group)  
 ORGANISM Oryza sativa (japonica cultivar-group)  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
 Ehrhartoideae; Oryzeae; Oryza.  
 1 (bases 1 to 683)  
 Sasaki,T. and Yamamoto,K.  
 Rice cDNA from panicle at ripening stage (2000)  
 Unpublished  
 Contact: Takuji Sasaki  
 National Institute of Agrobiological Resources  
 Rice Genome Research Program, Kannondai 2-1-2, Tsukuba, Ibaraki  
 305-8602, Japan  
 Tel: 81-298-38-7441  
 Fax: 81-298-38-7468  
 Email: tsasaki@abr.affrc.go.jp, URL:http://rgp.dna.affrc.go.jp/  
 PROJECT="RGP".

## FEATURES

Location/Qualifiers  
 1..683

/organism="Oryza sativa (japonica cultivar-group)"  
 /mol\_type="mRNA"  
 /cultivar="Nipponbare"  
 /db\_xref="taxon:39947"

/clone="E11772"  
 /dev\_stage="ripening stage"

/clone\_lib="Rice panicle at ripening stage"  
 /note="Organ: panicle; Rice cDNA from panicle at ripening  
 stage"

BASE COUNT 146 a 179 c 242 g 115 t 1 others  
 ORIGIN

## Alignment Scores:

Pred. No.: 8,31e-22 Length: 683  
 Score: 432.50 Matches: 104  
 Percent Similarity: 63.27% Conservative: 20  
 Best Local Similarity: 53.06% Mismatches: 44  
 Query Match: 37.67% Indels: 29  
 DB: 9 Gaps: 8

US-10-053-410-4 (1-206) x AUI64031 (1-683)

QY 16 AlaLeuValAlaValAlaValCysGlnGlyGluValGluArgGlnArgLeuArgAspLeu 35  
 Db 4 GCCATGGTGGCCATCTCCGCGCGGAGAGTTCAGGATCGAAGATGAGTTCAGGGACAGC 63  
 QY 36 GlnCysTrpGlnGlnValGlnGluSerProLeuAspAlaCysArgGlnValLeuAspArg 55  
 Db 64 CAGTCCAGCGGAGGTGCAGGACGCCGCTGGACGCGTCCGCGAGGTCTCGACCGG 123  
 QY 56 GlnLeuThrGly-GlyGlyGlyGlyValGlyValGlyProPheArgTrpGlyThrGlyLe 75  
 Db 124 CAGCTACCGCGCGGAGAGGTTCAGCGCGATGTTCCCGCGCGCGCGCGCT 182  
 QY 75 uArgMetArgCysCysGlnGlnLeuGlnAspValSerArgGluCysArgCysAlaAlaI 95  
 Db 183 CGCGATGTCAGTCTGCCAGCAGCTGCAGGACGTGAGCGCGAGTCCGCTGCGCGCCAT 242  
 QY 95 eArgSerMetValArgGlyTyrGluGluAlaMetPro----ProLeuGluGlyTrpTr 114

```
|||||
243 CCSCCGGATGCTGAGGAGCTACAGAGAGAGATGCCCGTCCCTCGAGCAGAGCTGGTC 302
QY 114 pProTTPGlyArgGlnGlnProProGlnGlnGlyGlyGlyGlyGlyGlyGlyTy 134
Db 303 GTCTGCTGCTCGAGTACTAC-----GGCGCGAGGG----- 336
QY 134 rTyTyTyProCysSerArgProGlyGlnGlyTyTyTyTyTyTyTyTyTyTyTy 154
Db 337 -----TCGTCGTCGAGCAGGGGTAC---TACGGCGAGGGGTCTCGGAGGA 380
QY 154 nMetTyTy-----ProProCysArgProGlyTyTyTyTyTyTyTyTyTyTy 171
Db 381 GGCTACTACGGCGAGCAGCAGCAGCGCGGGATGACC----- 420
QY 171 yArgValArgLeuThrLysAlaArgGluTyTyAlaAlaGlyLeuProMetMetCysArgLe 191
Db 421 -CGCGTGAGGCTGACCCAGGCGAGGAGGATACGCGCGGCGAGCTGCCGTCGATGTCGGGGT 479
QY 191 userGluProGlnGluCysSerIlePheSerGlyGlyAspGlnTy 206
Db 480 T---GAGCCCGCAGAGTCAGCATCTTCGCCCGCGGC---CAGTAC 519
```

```
RESULT 15
AI670620/c
LOCUS AI670620 389 bp mRNA linear EST 02-FEB-2000
DEFINITION 605035H03.x1 605 - EndospERM cDNA library from Schmidt lab Zea mays
cDNA, mRNA sequence.
ACCESSION AI670620
VERSION AI670620.1 GI:4837530
KEYWORDS EST.
SOURCE Zea mays
ORGANISM Zea mays
REFERENCE 1. (bases 1 to 389)
AUTHORS Walbot,V.
TITLE Maize ESTs from various cDNA libraries sequenced at Stanford
JOURNAL Unpublished
COMMENT Contact: Walbot V
Department of Biological Sciences
Stanford University
855 California Ave, Palo Alto, CA 94304, USA
Tel: 650 723 2227
Fax: 650 725 8221
Email: walbot@stanford.edu
Plate: 605035 row: H column: 03.
FEATURES
source
1. .389
/organism="Zea mays"
/mol_type="mRNA"
/cultivar="Ohio43"
/db_xref="taxon:4577"
/tissue_type="nucellar, embryo, and endospERM"
/dev_stage="10-14 days post-pollination"
/lab_host="DH5 (alpha)"
/clone_lib="605 - EndospERM cDNA library from Schmidt lab"
/notes="Organ: Kernel; Vector: PAD-GAL4-2; Site: 1; EcoRI;
Site 2: XhoI; Kernel endospERM cDNA library from Schmidt
lab"
```

```
BASE COUNT 67 a 136 c 100 g 85 t 1 others
ORIGIN

Alignment Scores:
Pred. No.: 6e-22 Length: 389
Score: 431.00 Matches: 76
Percent Similarity: 96.25% Conservative: 1
Best Local Similarity: 95.00% Mismatches: 3
Query Match: 37.54% Indels: 0
DB: 9 Gaps: 0
```

```
US-10-053-410-4 (1-206) x AI670620 (1-389)
QY 127 GlyGlyGlyGlyGlnGlyGlyTyTyTyTyProCysSerArgProGlyGlyGlyTyGly 146
Db 387 GGAGGAGGAGGACAGGGGGGCTACTACTACCCCTTCAGCCCGCCAGGAGGATGCCGC 328
QY 147 TyTyTyGlnGlyGlyGlnArgGlnMetTyTyProProCysArgProGlyThrThrGlyGly 166
Db 327 TACGGTCAGGGTGGCCAGCGGCAGATGATATCCACCTGTCTGCCCGGACACCGCGCGGC 268
QY 167 GlyProArgIleGlyArgValArgLeuThrLysAlaArgGluTyTyAlaAlaGlyLeuPro 186
Db 267 GGGCCCAAGGATCGCGCGGTGAGGCTTNCGAAGGCCCGGGAGTACGCCCGGGTTGCCG 208
QY 187 MetMetCysArgLeuSerGluProGlnGluCysSerIlePheSerGlyGlyAspGlnTy 206
Db 207 ATGATGTGCCGCTGTCTCGAGGCCCGCAGGAGTGCAGGCTCTTCTCCGGCGCGGACGAGTAC 148
```

Search completed: November 29, 2003, 18:49:18  
Job time : 2052 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: November 29, 2003, 17:10:55 ; Search time 62 Seconds  
(without alignments)

1466.531 Million cell updates/sec

Title: US-10-053-410-4

Perfect score: 1148

Sequence: 1 MAKIAAAAAAALCFALVAV.....MMCRSLSEPQCSIFSGGQY 206

Scoring table: BLOSUM62

Xgapop 10.0 ; Xgapext 0.5

Fgapop 10.0 ; Fgapext 0.5

Delop 6.0 ; Delext 7.0

Searched: 569978 seqs, 220691566 residues

Total number of hits satisfying chosen parameters: 1139956

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters: -DBV=xlh

-Q/cgn2\_1/USPTO\_spool/US10053410/runat\_28112003.140757.27908/app\_query.fasta\_1.391  
-DB-issued Patents NA -Qfmt=fastp -SUFIX=p2n.rni -MINMATCH=0.1 -LOOPCL=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blom62 -TRANS=human40.cdi  
-LIST=45 -DOCALLIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US10053410 @CGN 1.1.56 @runat\_28112003.140757.27908 -NCPU=6 -ICPU=3  
-NO\_WMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -DSFLOCK=100 -LONGLOG  
-NO\_TIMEOUT=120 -WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Issued Patents NA:  
1: /cgn2\_6/ptodata/2/ina/5A COMB.seq:\*  
2: /cgn2\_6/ptodata/2/ina/5B COMB.seq:\*  
3: /cgn2\_6/ptodata/2/ina/6A COMB.seq:\*  
4: /cgn2\_6/ptodata/2/ina/6B COMB.seq:\*  
5: /cgn2\_6/ptodata/2/ina/PTCUS COMB.seq:\*  
6: /cgn2\_6/ptodata/2/ina/backfileseq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	281.5	24.5	2296	US-08-899-336-2	Sequence 2, Appli
2	130.5	11.4	477	US-08-706-391B-6	Sequence 6, Appli
3	124	10.8	554	US-09-252-991A-8810	Sequence 8810, Ap
4	124	10.8	1233	US-09-252-991A-9241	Sequence 9241, Ap
5	124	10.8	1848	US-09-252-991A-9071	Sequence 9071, Ap
6	124	10.8	1980	US-09-252-991A-8712	Sequence 8712, Ap
7	123	10.7	2824	US-09-010-928B-3	Sequence 3, Appli
8	118.5	10.3	5822	US-08-899-595-4	Sequence 4, Appli
9	118.5	10.3	5822	US-08-899-595-5	Sequence 5, Appli
10	117	10.2	1251	US-09-252-991A-2647	Sequence 2647, Ap
11	117	10.2	1323	US-09-252-991A-3022	Sequence 3022, Ap
12	117	10.2	1371	US-09-252-991A-2846	Sequence 2846, Ap

c 13	117	10.2	4403765	3	US-09-103-840A-2	Sequence 2, Appli
c 14	117	10.2	4411529	3	US-09-103-840A-1	Sequence 4, Appli
c 15	116.5	10.1	2889	1	US-08-537-002A-4	Sequence 4, Appli
c 16	116.5	10.1	2889	3	US-08-863-010-4	Sequence 4, Appli
c 17	116.5	10.1	2889	3	US-09-024-429-4	Sequence 5, Appli
c 18	116.5	10.1	3600	1	US-08-537-002A-5	Sequence 5, Appli
c 19	116.5	10.1	3600	3	US-08-863-010-5	Sequence 5, Appli
c 20	116.5	10.1	3600	3	US-09-024-429-5	Sequence 2, Appli
c 21	116	10.1	2872	4	US-09-327-487A-2	Sequence 1, Appli
c 22	116	10.1	47981	4	US-09-679-279-1	Sequence 10678, A
c 23	115	10.0	927	4	US-09-252-991A-10678	Sequence 10762, A
c 24	115	10.0	1272	4	US-09-252-991A-10762	Sequence 10284, A
c 25	115	10.0	1323	4	US-09-252-991A-10284	Sequence 10396, A
c 26	115	10.0	1425	1	US-07-668-648-3	Sequence 3, Appli
c 27	115	10.0	1779	1	US-08-429-998-3	Sequence 3, Appli
c 28	115	10.0	1779	2	US-08-431-333-3	Sequence 3, Appli
c 29	115	10.0	1779	5	PCT-US91-02321-3	Sequence 3, Appli
c 30	115	10.0	1779	5	PCT-US91-02321-3	Sequence 3, Appli
c 31	115	10.0	2152	4	US-09-016-434-1305	Sequence 1305, Ap
c 32	114.5	10.0	1907	4	US-09-443-184-40	Sequence 40, Appli
c 33	113.5	9.9	1299	4	US-09-252-991A-5223	Sequence 5223, Ap
c 34	113.5	9.9	1806	4	US-09-252-991A-5253	Sequence 5318, Ap
c 35	113.5	9.9	1929	4	US-09-252-991A-5318	Sequence 3, Appli
c 36	113.5	9.9	5115	3	US-08-348-518C-3	Sequence 3, Appli
c 37	113.5	9.9	5115	3	US-08-476-509B-3	Sequence 3, Appli
c 38	113	9.8	44377	2	US-08-804-198-1	Sequence 7, Appli
c 39	113	9.8	44377	2	US-08-804-198-1	Sequence 1, Appli
c 40	112.5	9.8	2787	4	US-09-254-594-2	Sequence 2, Appli
c 41	112.5	9.8	3195	4	US-09-254-594-1	Sequence 1, Appli
c 42	112	9.8	1410	4	US-09-506-859-14	Sequence 14, Appli
c 43	112	9.8	1410	5	PCT-US95-15428-14	Sequence 14, Appli
c 44	112	9.8	1817	3	US-08-564-496C-14	Sequence 14, Appli
c 45	112	9.8	2092	3	US-09-318-978-1	Sequence 1, Appli

#### ALIGNMENTS

#### RESULT 1

US-08-899-336-2

Sequence 2, Application US/08899336

Patent No. 5955649

GENERAL INFORMATION:

APPLICANT: HIROTA, NAHICO

APPLICANT: KIHARA, MAKOTO

APPLICANT: KURODA, HISAO

APPLICANT: ITO, KAZUTOSHI

TITLE OF INVENTION: GENE EXPRESSION REGULATORY DNA,

TITLE OF INVENTION: EXPRESSION CASSETTE, EXPRESSION VECTOR AND TRANSGENIC

TITLE OF INVENTION: PLANT

NUMBER OF SEQUENCES: 8

CORRESPONDENCE ADDRESS:

ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MATER & NEUSTADT,

ADDRESSEE: P.C.

STREET: 1755 SOUTH JEFFERSON DAVIS HIGHWAY, SUITE 400

CITY: ARLINGTON

STATE: VIRGINIA

COUNTRY: U.S.A.

ZIP: 22202

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent In Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/899,336

FILING DATE: 23-JUL-1997

CLASSIFICATION: 800

PRIOR APPLICATION DATA:

APPLICATION NUMBER: JP HEI 8-193433

FILING DATE: 23-JUL-1996

ATTORNEY/AGENT INFORMATION:

NAME: OBLON, NORMAN F.

REGISTRATION NUMBER: 24,618

issued 7/29/99  
102(6) on "46"

REFERENCE/DOCKET NUMBER: 2589-0061-0

TELECOMMUNICATION INFORMATION:

TELEPHONE: (703) 413-3000

TELEFAX: (703) 413-2220

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 2296 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)

US-08-899-336-2

#### Alignment Scores:

Pred. No.:	9,598-14	Length:	2296
Score:	281.50	Matches:	79
Percent Similarity:	46.12%	Conservative:	22
Best Local Similarity:	36.07%	Mismatches:	51
Query Match:	24.52%	Indels:	67
DB:	2	Gaps:	11

US-10-053-410-4 (1-206) x US-08-899-336-2 (1-2296)

```
QY 1 MetAlaLysIleAlaAlaAlaAlaAlaAlaAlaLeuValAlaVal 20
DB 37 ATGGCTAAGCGGTGGTCTCTTTGTGGCGGTATC-----GTCGCCCTCGTGGCTCTC 90
QY 21 AlaValCysGlnGlyGluValGlu-----ArgGlnArgLeuArgAsp 34
DB 91 ACCACCGCTGAACGTGAGATCAATGGAACAACATTTCCTTGATAGCGCTCTAGGCAG 150
QY 35 LeuGlnCysTrpGlnGluValGlnGluSerProLeuAspAlaCysArgGlnValLeuAsp 54
DB 151 CTACAGTGTGAGCGGAGCTCCAGGAGAGCTCGCTCGAGCGGTGCGGGCGGTGCGAC 210
QY 55 ArgGlnLeuThrGlyGlyGlyGlyGlyGlyValGlyValGlyProPheArgTrpGlyThrGly 74
DB 211 CAACAGCTG-----GTTGGCAGCTGCATGGAGCAGCGGG 246
QY 75 LeuArgMetArgCysCysGlnGlnLeuGlnAspValSerArgGluCysArgCysAlaAla 94
DB 247 CTCAGATGCAAGTGTCCAGCAGAGCTTCGGGAGCTCAGCCCCGAGTGCCTGCGCG 306
QY 95 IleArgSerMetValArgGlyTrpGluGlu-----104
DB 307 CTCAGCCAGTGTGTAGGCAATACGACGACGACAAACCGAGGTGCATCCAAAGGAGGATCC 366
QY 105 -----AlaMetProProLeuGluLys---GlyTrpTrp-----114
DB 367 TTCTACCGCGGGCGGACCGCCACCGCTGTCAGCAAGGAGGATGGTGGGGAACCTCTGTA 426
QY 115 -----ProTrpArgGlnGln-----120
DB 427 AATGGTACTACCCAGACCAAACTTCTTCGCAACAGTCATGGCAAGGGCAACAAGGGTAC 486
QY 121 -----GlnProProGlnGlyGlyGlyGlyGlnGlyGly 133
DB 487 CACCAAGCGTAACTTCTTCCAGCAGCCAGGCAAGGGCGCAGGGTCTTACCAGGT 546
QY 134 TyrTyrTyrProCysSerArgProGlyGlyGlyTyrGlyTyrGlyGlnGlyGlnArg 153
DB 547 TCNACTTTCGCG---CAGCAGCCAGGACAAAGGACAA-----CAACAGGACAGAGG 594
QY 154 Gln-----MetTyrPro-----ProCysArgProGlyThrThrGlyGly 166
DB 595 CAGCCATGGTCTTATCAAGTGCAACTTTCCTCCACCAACAGCCAGGCGCAAGGCAAGG 651
```

#### RESULT 2

US-08-706-391B-6

Sequence 6, Application US/08706391B

Patent No. 6174725

GENERAL INFORMATION:

APPLICANT: ANDERSON, OLIN D

TITLE OF INVENTION: ALTERING DOUGH VISCOELASTICITY WITH

MODIFIED GLUTENINS

NUMBER OF SEQUENCES: 11

CORRESPONDENCE ADDRESS:

ADDRESSEE: WEST, AGRICULTURAL RESEARCH SERVICE, PACIFIC

WEST

STREET: 800 BUCHANAN STREET

CITY: ALBANY

STATE: CALIFORNIA

COUNTRY: USA

ZIP: 94710

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent in Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/706,391B

FILING DATE: 30-Aug-1996

CLASSIFICATION: <Unknown>

ATTORNEY/AGENT INFORMATION:

NAME: CONNOR, MARGARET A

REGISTRATION NUMBER: 30,043

REFERENCE/DOCKET NUMBER: 0235.95/USDA96-001

TELECOMMUNICATION INFORMATION:

TELEPHONE: (510) 559-6067

TELEFAX: (510) 559-5777

INFORMATION FOR SEQ ID NO: 6:

SEQUENCE CHARACTERISTICS:

LENGTH: 477 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: linear

MOLECULE TYPE: CDNA

SEQUENCE DESCRIPTION: SEQ ID NO: 6:

US-08-706-391B-6

#### Alignment Scores:

Pred. No.:	0,0166	Length:	477
Score:	130.50	Matches:	55
Percent Similarity:	37.62%	Conservative:	21
Best Local Similarity:	27.23%	Mismatches:	53
Query Match:	11.37%	Indels:	73
DB:	3	Gaps:	11

US-10-053-410-4 (1-206) x US-08-706-391B-6 (1-477)

```
QY 24 GlnGlyGluValGluArgGlnArgLeuArgAspLeuGlnCys-----TrpGlnGlu 40
DB 4 GAAGGTGAAGCTCTGAACAG-----CTGCAGTGCATCGCGAAGTGCAGGAA 51
QY 41 ValGlnGluSerProLeuAspAlaCysArgGlnValLeuAspArgGlnLeuThrGlyGly 60
DB 52 CTGCAGGACGCGTAACTGAAAGCTTCCAGCAGAGTTATGGAC-----93
QY 61 GlyGlyGlyGlyValGlyProPheArgTrpGlyLeuArgMetArgCysCys 80
DB 93 -----93
QY 81 GlnGlnLeuGlnAspValSerArgGluCysArgCysAlaAlaIleArgSerMetValArg 100
DB 94 CAGCAGCTGCGTGACATCTCTCCGAATGCCACCGGTTGTTGTTCTCCGCTTGTCTG 153
QY 101 GlyTyrGluGlu-----AlaMetProProLeuGluLysGlyTyrTrpProTrpGlyArg 118
DB 154 CAGTACGACACACAGCATCGTTGTTCCGCG-----AAAGGTGGTACCTTCTATCCGGGT 207
QY 119 GlnGlnGlnProProGln-----GlyGlyGlyGlyGlnGly 132
DB 208 GAACACCACTCCCGCCGACGACGCTGACGAGCAGCTATCTTCTGGGGCATCCCGCTCTG 267
QY 133 GlyTyrTyrTyrPro-----CysSerArgProGlyGluGlyTyrGlyTyrGlyGln 149
```

Db		268	AAGCGTACTACCCGCTGTGTACTTGTCGCCGAG-----	300
Qy		150	GlyGlyGlnArgGlnMetTy <sup>r</sup> Pro-----ProCysArgProGlyThrTh <sup>r</sup> ---	164
Db		301	-----CAGGTTCTTACTACCGGGTCAGGCTTCCGCAGCGTTCTAGTCTTCTTAC	354
Qy		165	-----GlyGlyGlyProArgIleGlyArgValArgLeuThrLy <sup>s</sup> AlaA <sup>r</sup> gGluTy <sup>r</sup> Ala	182
Db		355	CACGTTTCTGTGTAACACACGAGCTGCATCTCTGAAGTTGCTAAAGCTCAGCAGCTGGCT	414
Qy		183	AlaGlyLeuProMetMetCysArgLeuSerGluProGlnGluCysSerIlePheSerGly	202
Db		415	GCGCAGCTGCCGGCTATGTGCCGCTG-----GAAGGT	447
Qy		203	GlyAsp	204
Db		448	GGTGAC	453

### RESULT 3

```

RESULI 3
US-09-252-991A-8810/c
; Sequence 8810, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 8810
; LENGTH: 654
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-8810

```

Alignment Scores:		
Pred. No.:	0.0794	Length:
Score:	124.00	Matches:
Percent Similarity:	43.72%	Conservative:
Best Local Similarity:	32.66%	Mismatches:
Query Match:	10.80%	Indels:
DB:	4	Gaps:
		12
		654

US-10-053-410-4 (1-206) x US-09-252-991A-8810 (1-654)

[illegible][illegible]

## RESULT 4

```

RES001.4
US-09-252-991A-9241
; Sequence 9241, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 9241
; LENGTH: 1233
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-9241

```

Alignment Scores:		
Pred. No.:	0.168	Length: 1233
Score:	124.00	Matches: 65
Percent Similarity:	43.72%	Conservative: 22
Best Local Similarity:	32.66%	Mismatches: 75
Query Match:	10.80%	Indels: 38
DB:	4	Gaps: 12

US-10-053-410-4 (1-206) x US-09-252-991A-9241 (1-1233)

Qy	1	MetAlaLysIleLeaAlaAlaAlaAlaLeuCysPheAlaAlaLeuValaVal	20
Db	86	CTGCCCAGACTGGAGTCCCGAGGCCTGCCGTTCCTGATTGCGGCCCTG-----	136
Qy	21'	AlaValCysGlnGlyGlu-----ValGluArgGlnArgLeuArgAspLeuGln	36
Db	137	---::: ::: --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---	187
Qy	37	CysTrpGlnGluValGlnGluSerProLeuAspAlaCysArgGlnValLeuAspArgGln	56
Db	188	GCCTTGGAGGAA-----GACGATGTGTGATGCGTCTGTGCGGCCCTTGCG- CGATCC	240
Qy	57	LeuThrGlyGlyGlyGlyGlyGlyValGlyProPheArg---TrpGlyThrGlyLeu	75
Db	241	GGTCCCGCGGTGGCCGATGCCGGCGCACAGCCTTGGCGNACTCAAGAACCGGCCGCG	300
Qy	76	ArgMetArgCysCysGlnGlnLeuGlnAspValSerArgGluCysArgCysAlaAlaIle	95
Db	301	CGGCGCGCGCTGTGCC-----GTGGCTCGGCATGCCGATGCCTTCGTCCG	348
Qy	96	ArgSerMetValArgGlyTyrGluGluAlaMetProProLeuGluLysGlyTrpTrpPro	115
Db	349	CGCCAGCGGTATTGGCTGGTTCGCGCACTCGCGCTGGAGAGACGCGCGCGCGCCCT	408

```
Qy 116 TrpGlyArgGlnGlnGlnPro-----ProProGlnGlyGlyGlyGlyGln 131
Db 409 ---GGCGCCCTCGGCATCCGACGCGCGGTGCGCGGAGCGGTGCGGTGCTCGG 465
Qy 132 GlyGlyTyrTyrProCysSerArgProGlyGluGlyTyrGlyTyrGlnGlyGly 151
Db 466 -----CTGGCTCGGCACACGCGCGCTGCGCGA-----ACTGGCCAGGCTGGC 510
Qy 152 GlnArgGlnMetTyrProProCysArgProGlyThrThrGlyGlyGlyProArgIleGly 171
Db 511 CAGCGCGGA-----TGTCCACCGGAGGTACCGCGCGCGCCACCG----- 552
Qy 172 ArgValArg-LeuThrLysAlaArgGluTyrAlaAlaGlyLeuProMetMetCys 189
Db 553 -----CGCCTCGGCCTGAGCGCGAG---GCGACGGTGTGCGCGCGCTCTGC 598

RESULT 5
US-09-252-991A-9071
; Sequence 9071, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR FILING DATE: 1998-02-18
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 9071
; LENGTH: 1848
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-9071

Alignment Scores:
Pred. No.: 0.271 Length: 1848
Score: 124.00 Matches: 65
Percent Similarity: 43.72% Conservative: 22
Best Local Similarity: 32.66% Mismatches: 75
Query Match: 10.80% Indels: 38
DB: 4 Gaps: 12

US-10-053-410-4 (1-206) x US-09-252-991A-9071 (1-1848)
Qy 1 MetAlaLysIleAlaAlaAlaAlaAlaLeuCysPheAlaAlaLeuValAlaVal 20
Db 979 CTGGCCCACTGGAGCTGCGGAGCGCTGCGGTGCTGATTGCGGCCCTG----- 1029
Qy 21 AlaValCysGlnGlyGlu-----ValGluArgGlnArgLeuArgAspLeuGln 36
Db 1030 -----CGCGCGATCCGACCCCGCGGTGCGCGCGAGCGCGCGCTGCTGGAG 1080
Qy 37 CysTrpGlnGluValGlnGluSerProLeuAspAlaCysArgGlnValLeuAspArgGln 56
Db 1081 GCCTGGGAGGAA-----GACGCAAGTGTGATCGCTGTGCGGCCCTGGC-CGATCC 1133
Qy 57 LeuThrGlyGlyGlyGlyGlyValGlyValGlyProPheArg---TrpGlyThrGlyLeu 75
Db 1134 GTGCGCGCGGTGCCATGCGCGCGAGAGCTTGGCGAACTCAAGAACCGCGCGC 1193
Qy 76 ArgMetArgCysGlnGlnLeuGlnAspValSerArgGluCysArgCysAlaAlaIle 95
Db 1194 CGGCGCGCGCTGCTGCC-----GTGGCTCGGCGATCGCGATGCTTCGTCGCG 1241
Qy 96 ArgSerMetValArgGlyTyrGluGluAlaMetProLeuGluLysGlyTyrTrpPro 115
Db 1242 CGCCAGCGTATTGCGGTGCGCGAACTGCGCGCTGAGGAGAGCGCGCGCGCCCT 1301
Qy 116 TrpGlyArgGlnGlnGlnPro-----ProProGlnGlyGlyGlyGlyGln 131
```

```
Db 1302 ---GGCGCCCTCGGCATCCGACGCGCGGTGCGCGGAGCGGTGCGGTGCTCGG 1358
Qy 132 GlyGlyTyrTyrProCysSerArgProGlyGluGlyTyrGlyTyrGlnGlyGly 151
Db 1359 -----CTGGCTCGGCACACGCGCGCTGCGCGA-----ACTGGCCAGGCTGGC 1403
Qy 152 GlnArgGlnMetTyrProProCysArgProGlyThrThrGlyGlyGlyProArgIleGly 171
Db 1404 CAGCGCGGA-----TGTCCACCGGAGGTACCGCGCGCGCCACCG----- 1445
Qy 172 ArgValArg-LeuThrLysAlaArgGluTyrAlaAlaGlyLeuProMetMetCys 189
Db 1446 -----CGCCTCGGCCTGAGCGCGAG---GCGACGGTGTGCGCGCGCTCTGC 1491

RESULT 6
US-09-252-991A-8712/c
; Sequence 8712, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR FILING DATE: 1998-02-18
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 8712
; LENGTH: 1980
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-8712

Alignment Scores:
Pred. No.: 0.294 Length: 1980
Score: 124.00 Matches: 65
Percent Similarity: 43.72% Conservative: 22
Best Local Similarity: 32.66% Mismatches: 75
Query Match: 10.80% Indels: 38
DB: 4 Gaps: 12

US-10-053-410-4 (1-206) x US-09-252-991A-8712 (1-1980)
Qy 1 MetAlaLysIleAlaAlaAlaAlaAlaAlaLeuCysPheAlaAlaLeuValAlaVal 20
Db 942 CTGGCCCACTGGAGCTGCGGAGCGCTGCGGTGCTGATTGCGGCCCTG----- 892
Qy 21 AlaValCysGlnGlyGlu-----ValGluArgGlnArgLeuArgAspLeuGln 36
Db 891 -----CGCGCGATCCGACCCCGCGGTGCGCGCGAGCGCGCGCTGCTGGAG 841
Qy 37 CysTrpGlnGluValGlnGluSerProLeuAspAlaCysArgGlnValLeuAspArgGln 56
Db 840 GCCTGGGAGGAA-----GACGCAAGTGTGATGCGCTGTGCGGCCCTGGC-CGATCC 788
Qy 57 LeuThrGlyGlyGlyGlyGlyValGlyValGlyProPheArg---TrpGlyThrGlyLeu 75
Db 787 GTGCGCGCGGTGCCATGCGCGCGAGAGCTTGGCGAACTCAAGAACCGCGCGC 728
Qy 76 ArgMetArgCysGlnGlnLeuGlnAspValSerArgGluCysArgCysAlaAlaIle 95
Db 727 CGGCGCGCGCTGCTGCC-----GTGGCTCGGCGATCGCGATGCTTCGTCGCG 680
Qy 96 ArgSerMetValArgGlyTyrGluGluAlaMetProLeuGluLysGlyTyrTrpPro 115
Db 679 CGCCAGCGTATTGCGGTGCGCGAACTGCGCGCTGAGGAGAGCGCGCGCGCCCT 620
Qy 116 TrpGlyArgGlnGlnGlnPro-----ProProGlnGlyGlyGlyGlyGln 131
```

Db 619 -----GGCCGCCCTCGGCATCCGACGCGCGGTGCGCCCGAGCGGTGCGGTGCTCGG 563  
QY 132 GlyGlyTyrTyrProCysSerArgProGlyGluGlyTyrGlyTyrGlyGlnGlyGly 151  
Db 562 -----CTGGCTCCGCGACCGCGCGCTGCGCGA-----ACTGCGCAGCGCTGCG 518  
QY 152 GlnArgGlnMetTyrProProCysArgProGlyThrGlyGlyProArgGlyGly 171  
Db 517 CAGCGCGCGA-----TGTGACCGGAGGTACCGCGCGCGCGCCACCGG----- 476  
QY 172 ArgValArg-LeuThrLysAlaArgGlyTyrAlaAlaGlyLeuProMetMetCys 189  
Db 475 -----CGCCCTCGGCTGAGCGCGAG---GCGACGGTGTGCGCGCGCTGCTGC 430

## RESULT 7

US-09-010-928B-3  
; Sequence 3, Application US/09010928B  
; Patent No. 5994099  
; GENERAL INFORMATION:  
; APPLICANT: Lewis, Randolph V  
; APPLICANT: Hayashi, Cheryl Y  
; TITLE OF INVENTION: EXTREMELY ELASTIC SPIDER SILK PROTEIN AND DNA  
; TITLE OF INVENTION: CODING THEREOF  
; NUMBER OF SEQUENCES: 29  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: BIRCH, STEWART, KOLASCH & BIRCH  
; STREET: 8110 GATEHOUSE RD. SUITE 500E  
; CITY: FALLS CHURCH  
; STATE: VIRGINIA  
; COUNTRY: UNITED STATES OF AMERICA  
; ZIP: 22042

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/010,928B  
FILING DATE: 22-JAN-1998  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Murphy Jr., Gerald M  
REGISTRATION NUMBER: 28977  
REFERENCE/DOCKET NUMBER: 1447-109P  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2824 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: not relevant  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
HYPOTHETICAL: NO  
FEATURE:  
NAME/KEY: -  
LOCATION: 1..2824  
OTHER INFORMATION: /note= "Flagelliform DNA sequence  
OTHER INFORMATION: taken from 3' region. Stop codon begins at position 2722."  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 1..2724  
US-09-010-928B-3

## Alignment Scores:

Pred. No.:	0.537	Length:	2824
Score:	123.00	Matches:	48
Percent Similarity:	38.93%	Conservative:	10
Best Local Similarity:	32.21%	Mismatches:	64
Query Match:	10.71%	Indels:	27
DB:	2	Gaps:	5

US-10-053-410-4 (1-206) x US-09-010-928B-3 (1-2824)

QY 32 LeuArgAspLeuGlnCysTrpGlnGluValGlnGluSerProLeuAspAlaCysArgGln 51

Db 715 ATAGAAGACTTGATATTAACATTCATGCGCTGATGCGCCGATA----- 759  
QY 52 ValLeuAspArgGlnLeuThr-----GlyGlyGlyGlyGlyGlyValGly 67  
Db 760 AGATTTTCAGAAAGTAATTAACATTAAGTGTGCTGAGAGTTCCGGACCCCGTGTCTGGA 819  
QY 68 ProPheArgTrpGlyThrGly--LeuArgMetArgCysGlnGlnLeuGlnAspValS 87  
Db 820 CCAGGAGGTGTAGGCGCTGTGCTCCGACCCAGAGGTGTAGGACCTGGAGGTTCTGGA 879  
QY 87 exArgGluCysArgCysAlaAlaLeuArgSerMetValArgGlyTyrGluGluAlaMetP 107  
Db 880 CCAGGAGGTGTAGGACCTGTGCTGCTGTCAGGAGCGCT----- 920  
QY 107 roProLeuGlnLysGlyTrpTrpProTrp-GlyArgGln-----GlnGlnPro 122  
Db 921 -----AGGACCTGGTGTGCTGCTGAGACCTTAGCGACCTGGCGGTCTTGGACCT 969  
QY 123 ProProGlnGlyGlyGlyGlnGlyGlyTyrTyrTyrProCysSerArgProGly 142  
Db 970 GGAGGTGCGAGCGGAGCTGGAGGACCTGGTGGACATACGGACCTGGTGGATCATATGA 1029  
QY 143 GluGlyTyrGlyTyrGlyGlnGlyGlnArgGlnMetTyrProProCysArgProGly 162  
Db 1030 CCTGCTGTGTTCCGGAGGACCCGCTGCTGCTGCGGACCATACCGACCTGGAGGTGAAGGA 1089  
QY 163 ThrThrGly--GlyGlyPro 168  
Db 1090 CCCGCTGTGCTGGCGGACCC 1110

## RESULT 8

US-08-899-595-4/c  
; Sequence 4, Application US/08899595  
; Patent No. 6111072  
; GENERAL INFORMATION:  
; APPLICANT: Narumiya, Shuh  
; APPLICANT: Takahashi, No. 6111072uaki  
; TITLE OF INVENTION: RHO TARGET PROTEIN HUMAN MDIA AND GENE  
; TITLE OF INVENTION: ENCODING SAME  
; NUMBER OF SEQUENCES: 14  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Foley & Lardner  
; STREET: 3000 K Street, N.W., Suite 500  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: USA  
; ZIP: 20007-5109  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/899,595  
FILING DATE: 24-JUL-1997  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 8-242701  
FILING DATE: 26-AUG-1996  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 9-90170  
FILING DATE: 25-MAR-1997  
ATTORNEY/AGENT INFORMATION:  
NAME: Stephen A. Bent  
REGISTRATION NUMBER: 29,768  
REFERENCE/DOCKET NUMBER: 049441/0112  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202)672-5300  
TELEFAX: (202)672-5399  
TELEX: 904136  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:



LENGTH: 5822 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA to mRNA  
ORIGINAL SOURCE:  
ORGANISM: Human  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 28..3972  
US-08-899-595-4

Alignment Scores:  
Pred. No.: 2.88 Length: 5822  
Score: 118.50 Matches: 62  
Percent Similarity: 43.09% Conservative: 16  
Best Local Similarity: 34.25% Mismatches: 63  
Query Match: 10.32% Indels: 42  
DB: 3 Gaps: 10

US-10-053-410-4 (1-206) x US-08-899-595-4 (1-5822)

QY 18 ValAlaValAlaValCysGlnGlyGluValGluArgGlnArgLeuArgAspLeuGlnCys 37  
DB 2383 GTGGAGGTGGAGGAATGCCAGGGCCTCCGGGAATGGAGGAGGTGGAGGATTCAGGAC 2324  
QY 38 TrpGlnGluValGlnGluSerPro-----LeuAspAlaCysArgGlnVal 52  
DB 2323 CACCAAGAGAGGGGAGGAGGAGGTGGCATTCTCTCCAGGCAAGGGAGGAGTG 2264  
QY 53 -----LeuAspArgGlnLeuThrGlyGlyGlyGlyGly----- 64  
DB 2263 GGGGGGAATTCAGCACTCCAGGC-AAAGGAGGTGGTGGGGGATTCAGCACTC 2205  
QY 65 ---GlyValGlyProPheArgTrpGlyThrGlyLeuArgMetArgCysGlnGlnLeu 83  
DB 2204 CCAGGCAAGAGG---AGTGGGGGATGGCAGTACCTCCAGGCAAGAGAGGTTGA 2149  
QY 84 GlnAspValSer-----ArgGluCysArgCysAlaAlaLeuArgSerMet 98  
DB 2148 AGGATGCCAACACCTCCAGCAAGAGGGGTGGAGGATGGTAGCATCCCGACACAA 2089  
QY 99 ValArgGlyTyr-GluGluAlaMetProProLeuGluLysGlyTrpTrpProTrpGlyAr 118  
DB 2088 AGAGGGGTGGAGATACGATACCTCCAGTAAGAAGG-----GGTGA 2041  
QY 118 GlnGlnGlnProProPro---GlnGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGly 137  
DB 2040 GGAGATGCAACACCCCGAGGCAAGAGGTGGAGGAGGAGGAGGA----- 1992  
QY 137 oCysSerArgProGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGly 153  
DB 1991 -----GGAGGAGGAGGAGGAGGTGGTACTATCCCGAGGAGGAGGTGGTGG 1945  
QY 153 GlnMetTyrPro-----ProCysArgProGlyThrThrGlyGlyGlyGlyGlyGly 171  
DB 1944 AATAATAGTCCAGATCCACCGTAAAGAGGGGAGGAGGAGGAGGAGGAGGAGGAGG 1885  
QY 171 Y 171  
DB 1884 A 1884

## RESULT 9

US-08-899-595-5  
Sequence 5, Application US/0889955  
Patent No. 6111072  
GENERAL INFORMATION:  
APPLICANT: Natumiya, Shuh  
APPLICANT: Takahashi, No. 6111072uaki  
TITLE OF INVENTION: RHO TARGET PROTEIN HUMAN MDIA AND GENE  
TITLE OF INVENTION: ENCODING SAME  
NUMBER OF SEQUENCES: 14  
CORRESPONDENCE ADDRESS:

ADDRESSEE: Foley & Lardner  
STREET: 3000 K Street, N.W., Suite 500  
CITY: Washington  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20007-5109  
COMPUTER READABLE FORM:  
MEDIUM TYPE: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/899,595  
FILING DATE: 24-JUL-1997  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 8-242701  
FILING DATE: 26-AUG-1996  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 9-90170  
FILING DATE: 25-MAR-1997  
ATTORNEY/AGENT INFORMATION:  
NAME: Stephen A. Bent  
REGISTRATION NUMBER: 29,768  
REFERENCE/DOCKET NUMBER: 049441/0112  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202)672-5300  
TELEFAX: (202)672-5399  
TELEX: 904136  
INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 5822 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA to mRNA  
ANTI-SENSE: YES  
ORIGINAL SOURCE:  
ORGANISM: Human  
US-08-899-595-5

Alignment Scores:  
Pred. No.: 2.88 Length: 5822  
Score: 118.50 Matches: 62  
Percent Similarity: 43.09% Conservative: 16  
Best Local Similarity: 34.25% Mismatches: 63  
Query Match: 10.32% Indels: 42  
DB: 3 Gaps: 10

US-10-053-410-4 (1-206) x US-08-899-595-5 (1-5822)

QY 18 ValAlaValAlaValCysGlnGlyGluValGluArgGlnArgLeuArgAspLeuGlnCys 37  
DB 3440 GTGGAGGTGGAGGAATGCCAGGGCCTCCGGGAATGGAGGAGGTGGAGGATTCAGGAC 3499  
QY 38 TrpGlnGluValGlnGluSerPro-----LeuAspAlaCysArgGlnVal 52  
DB 3500 CACCAAGAGAGGGGAGGAGGAGGTGGCATTCTCTCCAGGCAAGGGAGGAGTG 3559  
QY 53 -----LeuAspArgGlnLeuThrGlyGlyGlyGlyGlyGlyGlyGlyGlyGly 64  
DB 3560 GGGGGGAATTCAGCACTCCAGGC-AAAGGAGGTGGTGGGGGATTCAGCACTC 3618  
QY 65 ---GlyValGlyProPheArgTrpGlyThrGlyLeuArgMetArgCysGlnGlnLeu 83  
DB 3619 CCAGGCAAGAGGAG---AGTGGGGGATGGCAGTACCTCCAGGCAAGAGAGGAGTGA 3674  
QY 84 GlnAspValSer-----ArgGluCysArgCysAlaAlaLeuArgSerMet 98  
DB 3675 AGGATGCCAACACCTCCAGGCAAGAGGAGGAGGTGGAGGATGGTAGCATCCCGACACAA 3734  
QY 99 ValArgGlyTyr-GluGluAlaMetProProLeuGluLysGlyTrpTrpProTrpGlyAr 118







```
QY      64 GlyGlyValGly---ProPheArgTrpGlyThrGlyLeuArgMetArgCysGlnGln 82
Db      2476 GGAGGGGAGCGCACCCCGAGAGCGGACCCAGGGCTTCGGTCTCGTCAGGGCCCCCGG 2417
QY      83 LeuGlnAspValSerArgGluCysArgCysAlaAlaIleArgSerMetValArgGlyTyr 102
Db      2416 GAAGGAGCTGGGGTCCCGGGGGCTGCCCC-----GGAGGCGGGGAAGGCTTT 2366
QY     103 GluGluAlaMetProProLeuGluLys-----GlyTrpTrp 114
Db     2365 CCAGGAGCCGACCTCCAGGGCGCGAAGAGGTCTGCGGGCGGCCCGGGGGAGGCTTC 2306
QY     115 ProTrpGlyArg-----GlnGlnProProGlnGlyGly----- 127
Db     2305 CCGTGAGGGCGAGACCCCTTCTGGAGGCGCCCGCTCCAGAAAGAGGCCCTTCGGGCC 2246
QY     128 -----GlyGlyGlyGlnGlyGlyTyrTyrProCysSerArgProGlyGlu 143
Db     2245 GGAGAACCCAGGGAGGTCCAGGCGGGGAGGACCCGCTCCGTGCGGT---CCAGGCCCC 2189
QY     144 GlyTyrGlyTyrGlyGlnGlyGlyGlnArgGlnMetTyrProProCysArgProGlyThr 163
Db     2188 CGTCTTGGACGAGCGGAGCTGGACCCAGACCCCTCCCGCGCGAGTCCCG----- 2135
QY     164 ThrGlyGlyGlyProArgIleGlyArgValArgLeuThrLysAlaArgGluTyrAlaAla 183
Db     2134 GCCGGAGGAGTCCACGG-----CCTCGGGACCGGGACCCCGGGTGGC 2093
QY     184 Gly 184
Db     2092 GGC 2090
```

Search completed: November 29, 2003, 19:16:29  
Job time : 1629 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 29, 2003, 13:23:43 ; Search time 317 Seconds  
(without alignments)  
8089.799 Million cell updates/sec

Title: US-10-053-410-3  
Perfect score: 950  
Sequence: 1 aaaaaaacccctcgat.....aaaaaaaaaaaaaaaaaaaaa 950

Scoring table: IDENTITY NUC  
Gapop 10.0, Gapext 1.0

Searched: 2552756 seqs, 1349719017 residues  
Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : N Geneseq 19Jun03.\*  
1: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT.\*  
2: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT.\*  
3: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT.\*  
4: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT.\*  
5: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT.\*  
6: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT.\*  
7: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT.\*  
8: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT.\*  
9: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT.\*  
10: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT.\*  
11: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT.\*  
12: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT.\*  
13: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT.\*  
14: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT.\*  
15: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT.\*  
16: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT.\*  
17: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT.\*  
18: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT.\*  
19: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT.\*  
20: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT.\*  
21: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT.\*  
22: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT.\*  
23: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT.\*  
24: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT.\*  
25: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2003.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	91.2	9.6	2296	19 AAV20662	Barley D-hordein g
2	91.2	9.6	2380	22 AAF63380	Barley cDNA involv
3	91.2	9.6	2434	22 AAF63379	Barley cDNA involv
4	72.4	7.6	114955	20 AAX53491	Human adenosine A1
C 5	64.6	6.8	1614	22 AAH26499	Human low density
C 6	64.6	6.8	12425	22 AAH26495	Human low density
C 7	63.4	6.7	33957	22 AAA09686	HSV-2 immediate ea
8	63.4	6.7	154746	24 AAD25519	Human herpesvirus

C	9	63.4	6.7	154746	24	AAD25519	Human herpesvirus
10	63	6.6	88421	24	AAL40781	88421nt genomic DN	
11	62.4	6.6	1272	20	AA09010	Brn-3a polynucleot	
12	62.4	6.6	1272	21	AA29006	Human transcriptio	
C	13	61.6	6.5	114955	20	AAX53491	Human adenosine A1
14	60	6.3	1127	21	AA02477	Human colon cancer	
15	59.4	6.3	4524	20	AAV33912	Nucleotide sequenc	
C	16	59	6.2	3035	24	ABK09774	Human ovarian tumo
C	17	59	6.2	3056	24	ABK84722	Human cDNA differe
C	18	59	6.2	3056	25	ACA56555	Human signalling p
C	19	59	6.2	6695	22	AAK70340	Human immune/haema
C	20	59	6.2	6695	22	AAK82282	Human immune/haema
21	57.8	6.1	43411	24	ABQ88169	Human osteoblast d	
C	22	57.4	6.0	3169	22	AAI58822	Human polynucleoti
C	23	57.4	6.0	3203	22	AAI60608	Human polynucleoti
C	24	57.4	6.0	4190	15	AAQ55350	Sequence of the un
25	56.4	5.9	1556	18	AA794535	Maize cinnamoyl Co	
26	56	5.9	2549	25	ABZ24709	Human cell growth,	
27	56	5.9	2810	25	ABZ24708	Human cell growth,	
28	56	5.9	3147	23	AA85126	DNA encoding novel	
C	29	56	5.9	65140	22	AA17184	Streptomyces nous
C	30	56	5.9	125401	22	AA17186	Streptomyces nous
C	31	55.8	5.9	6225	20	AA55273	Human enzyme-relat
C	32	55.8	5.9	6225	21	AA20843	Human multiple tar
C	33	55.8	5.9	6225	21	AAA34721	Human adenosine re
C	34	55.8	5.9	35384	21	AA21436	Human enzyme-relat
C	35	55.6	5.9	594	24	ABQ43958	Oligonucleotide fo
C	36	55.6	5.9	594	24	ABQ43959	Oligonucleotide fo
C	37	55.6	5.9	5811	24	AB578655	S. macromyceticus
38	55.4	5.8	1000	21	AAA02484	Human colon cancer	
39	55.4	5.8	1698	23	AA84408	Human colon cancer	
40	55.4	5.8	2049	23	AA576210	DNA encoding novel	
C	41	55.4	5.8	2561	22	AAH26500	Rabbit low density
C	42	55.4	5.8	4403765	22	AA199683	Mycobacterium tube
C	43	55.2	5.8	1511	24	AB111009	Human breast cance
44	55.2	5.8	2668	22	AAK75409	Human immune/haema	
C	45	55	5.8	2687	21	AAZ30700	S.clavuligerus cla

ALIGNMENTS

RESULT 1  
AAV20662  
ID AAV20662 standard; DNA; 2296 BP.  
XX AAV20662;  
AC AAV20662;  
XX  
DT 23-JUN-1998 (first entry)  
XX  
DE Barley D-hordein gene 5'-terminal region.  
XX  
KW Barley; D-hordein; DPP3; promoter; gene expression; regulation;  
KW seed; structural gene; ds.  
XX  
OS Hordeum vulgare.  
XX  
PN WO9803655-A2.  
XX  
PD 29-JAN-1998.  
XX  
PF 22-JUL-1997; 97WO-JP02526.  
XX  
PR 23-JUL-1996; 96JP-0193433.  
XX (SAPB ) SAPPORO BREWERIES.  
XX  
PI Hirota N, Ito K, Kihara M, Kuroda H;  
XX  
DR WPI; 1998-120779/11.  
XX  
PT Gene expression regulatory DNA, expression cassettes and vectors -  
comprising promoter region from barley, Hordeum vulgare, D-hordein

PT gene, useful to control expression of desired gene e.g. to improve  
XX seeds  
PS Example 3; Page 25-26; 42pp; English.  
XX  
CC The present sequence represents the 5'-terminal region of the barley  
CC D-hordein gene. The present invention describes gene expression  
CC regulatory DNA which comprises a promoter region derived from the  
CC barley (Hordeum vulgare) D-hordein gene which allows expression of a  
CC desired structural gene, and a regulatory region regulating such  
CC expression. The introduction into plants of expression cassettes  
CC containing the gene expression regulatory DNA (either directly or via  
CC expression vectors) enables the expression of a gene within a plant cell  
CC e.g. barley to be controlled. The use of activating and suppressing  
CC regions in the regulatory DNA allows control of expression by e.g.  
CC tissue type or developmental stage, whilst the use of only an activating  
CC region maintains expression at a high level, providing an effective  
CC production means when recovery of the product of the structural gene is  
CC desired. The expression cassette/vector may be introduced into e.g.  
CC maturing seed endosperm tissue or regeneratable plant cells (e.g. derived  
CC from anthers) to improve seeds of barley or other plants, to produce  
CC gene products in seeds or to contribute to plant breeding programmes.  
CC The expression regulatory DNA can also be used in expression systems in  
CC vitro. GUS activity in barley protoplasts transfected with plasmid  
CC DpP3GUS2 comprising isolated D-hordein promoter region was 1.5 times  
CC higher than in those transfected with control pACTIF.  
XX  
SQ Sequence 2296 BP; 629 A; 708 C; 606 G; 353 T; 0 other;  
  
Query Match 9.6%; Score 91.2; DB 19; Length 2296;  
Best Local Similarity 61.4%; Pred. No. 2.6e-07;  
Matches 178; Conservative 0; Mismatches 88; Indels 24; Gaps 1;  
  
QY 207 AGGACCTCGAGTGTGGCAGGAGGTCCAGAGAGCCCGCTGACCGGTGCCCGCCAGGTC 266  
DB 145 AGGCAGCTACAGTGTGAGCGCGAGCTCCAGAGAGCTCGTCCAGGCGTCCCGCGGTC 204  
QY 267 CTCGACCGCAGCTACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 326  
DB 205 GTGGACCAACAGCTG-----GTTGGCCAGCTGCATGGAGC 240  
QY 327 ACCGGCTCCCGATGCGGTGCTGCCAGCAGCTCCAGGAGCTGAGCCGCGAGTCCCGCTGC 386  
DB 241 ACGGGGCTCCAGTGCAGTGTGCTGCCAGCAGCTTCGGGAGCTGAGCCCGCGCGCC 300  
QY 387 GCCGCATCCGAGCATGTGTACGGGCTACGAGAGGCCATGCCCGCTGGAGAAAGGC 446  
DB 301 GTCGCCCTCAGCCAGCTGCTGAGGCAATACGAGCAGCAAAACCGAGTGCCATCCAGGGA 360  
QY 447 TGGTGGCCATGGGGCGGCAGCAGCAGCCCGCCCGCGAGGAGGAGG 496  
DB 361 GGATCCTTCTACCGCGCGGAGCCCGCAGCGCGCTGCAGCAAGGAGATG 410  
  
RESULT 2  
AAF63380  
ID AAF63380 standard; cDNA; 2380 BP.  
XX  
AC AAF63380;  
XX  
DT 10-MAY-2001 (first entry)  
XX  
DE Barley cDNA involved in D-hordein production SEQ ID 2.  
XX  
KW Barley; gel protein; D-hordein; malt; wort filterability; extraction; ss.  
XX  
OS Hordeum vulgare.  
XX  
PN WO200111946-A1.  
XX  
PD 22-FEB-2001.  
XX  
PF 16-AUG-2000; 2000WO-JP05476.  
XX  
PA (SAPP) SAPPORO BREWERIES LTD.  
XX  
PI Hirota N, Kihara M, Ito K;

XX 16-AUG-1999; 99JP-0229696.  
XX (SAPP) SAPPORO BREWERIES LTD.  
XX  
PI Hirota N, Kihara M, Ito K;  
XX  
DR WPI; 2001-191587/19.  
XX  
PT Transformation of barley with a D-hordein expression regulator for  
PT production of barley with lowered gel protein content -  
XX  
PS Claim 4; Page 33-34; 40pp; Japanese.  
XX  
CC This invention relates to barley having a low gel protein content. The  
CC barley is transformed with a polynucleotide sequence which regulated the  
CC formation of D-hordein. Transformation results in the production of  
CC barley strains with improved malting properties such as wort  
CC filterability and efficiency of extraction. The present sequence  
CC represents cDNA involved in the regulation of D-hordein production.  
XX  
SQ Sequence 2380 BP; 642 A; 747 C; 622 G; 369 T; 0 other;  
  
Query Match 9.6%; Score 91.2; DB 22; Length 2380;  
Best Local Similarity 61.4%; Pred. No. 2.6e-07;  
Matches 178; Conservative 0; Mismatches 88; Indels 24; Gaps 1;  
  
QY 207 AGGACCTCGAGTGTGGCAGGAGGTCCAGGAGAGCCCGCTGACCGGTGCCCGCCAGGTC 266  
DB 132 AGGCAGCTACAGTGTGAGCGCGAGCTCCAGGAGAGCTCGTCCAGGCGGTGCCCGGTC 191  
QY 267 CTCGACCGCAGCTAACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 326  
DB 192 GTGGACCAACAGCTG-----GTTGGCCAGCTGCATGGAGC 227  
QY 327 ACCGGCTCCCGATGCGGTGCTGCCAGCAGCTCCAGGAGCTGAGCCGCGAGTCCCGCTGC 386  
DB 228 ACGGGGCTCCAGTGCAGTGTGCTGCCAGCAGCTTCGGGAGCTGAGCCCGCGCGCC 287  
QY 387 GCCGCATCCGAGCATGTGTACGGGCTACGAGGAGGCCATGCCCGCTGGAGAAAGGC 446  
DB 288 GTCGCCCTCAGCCAGCTGCTGAGGCAATACGAGCAGCAAAACCGAGTGCCATCCAGGGA 347  
QY 447 TGGTGGCCATGGGGCGGCAGCAGCAGCCCGCCCGCGAGGAGGAGG 496  
DB 348 GGATCCTTCTACCGCGCGGAGCCCGCAGCGCGCTGCAGCAAGGAGATG 397  
  
RESULT 3  
AAF63379  
ID AAF63379 standard; cDNA; 2434 BP.  
XX  
AC AAF63379;  
XX  
DT 10-MAY-2001 (first entry)  
XX  
DE Barley cDNA involved in D-hordein production SEQ ID 1.  
XX  
KW Barley; gel protein; D-hordein; malt; wort filterability; extraction; ss.  
XX  
OS Hordeum vulgare.  
XX  
PN WO200111946-A1.  
XX  
PD 22-FEB-2001.  
XX  
PF 16-AUG-2000; 2000WO-JP05476.  
XX  
PR 16-AUG-1999; 99JP-0229696.  
XX  
PA (SAPP) SAPPORO BREWERIES LTD.  
XX  
PI Hirota N, Kihara M, Ito K;





Db 108854 CGCCTGGCTCGCTGGGCGCCSNNNDNNGGGGGCGGCGGCGCTGGCTCGCTGGGCGC 108913  
QY 662 GTTGGCGATGATGCGCGGCTGCGAGAGCCCGAGGAGTGAGCATCTTCTCCGGCGCGCA 721  
Db 108914 CCSNNNDNNGGGCGGCGGCTGGCTCGCTGGGCGCCSNNNDNNGGGCGGCGCGC 108973  
QY 722 C 722  
Db 108974 C 108974

RESULT 5  
AAH26499/c  
ID AAH26499 standard; DNA; 1614 BP.  
XX  
AC AAH26499;  
XX  
DT 12-NOV-2001 (first entry)  
XX  
DE Human low density lipoprotein binding protein 2 (LBP-2) DNA.  
XX  
KW Low density lipoprotein binding protein 2; LBP-2; LDL; human;  
KW atherosclerosis; antiarteriosclerotic; gene therapy; diagnosis;  
KW vaccine; ds.  
XX  
OS Homo sapiens.  
XX  
PN WO200164874-A2.  
XX  
PD 07-SEP-2001.  
XX  
PF 28-FEB-2001; 2001WO-US063356.  
XX  
PR 02-MAR-2000; 2000US-0517849.  
PR 14-JUL-2000; 2000US-0616289.  
XX  
PA (BOST-) BOSTON HEART FOUND INC.  
XX  
PI Lees AM, Lees RS, Law SW, Arjona AA;  
XX  
XX WPI; 2001-565505/63.  
DR P-PSDB; AAB82806.  
XX  
PT New isolated low density lipoprotein binding polypeptide for treating,  
PT diagnosing and/or identifying therapeutic agents for atherosclerosis -  
XX  
PS Claim 7; Fig 7A; 143bp; English.

XX The present sequence is that of the coding region of the human  
CC gene (see also AAH26494) encoding novel human low density lipoprotein  
CC binding protein 2 (LBP-2, see AAB82806). The gene was isolated  
CC from a genomic DNA library using LBP-2 cDNA as probe. The LBP-2  
CC protein predicted from the present sequence differs from that  
CC predicted from a cDNA clone (see AAB82803) in that it contains an  
CC additional 321 amino acids at its N-terminus (the cDNA is a 5'  
CC truncation). LBP-2 nucleic acids are among claimed polynucleotides  
CC of the invention that encode novel polypeptides capable of binding  
CC to native and methylated LDL. Also claimed are isolated LBP  
CC polypeptides, and biologically active fragments and analogues of  
CC them, as well as expression vectors, cells and methods of producing  
CC the LBPs. Methods of determining if an animal is at risk for  
CC atherosclerosis, methods for evaluating an agent for use in treating  
CC atherosclerosis, and methods for treating a cell having an  
CC abnormality in structure or metabolism of LBP are claimed.  
CC Pharmaceutical compositions comprising an LBP polypeptide or nucleic  
CC acid, and vaccine compositions, are also claimed.  
XX  
SQ Sequence 1614 BP; 243 A; 638 C; 564 G; 169 T; 0 other;  
Query Match 6.8%; Score 64.6; DB 22; Length 1614;  
Best Local Similarity 46.4%; Pred. No. 0.012;  
Matches 211; Conservative 0; Mismatches 244; Indels 0; Gaps 0;

QY 108 GAGATGGCTAAGATCGCCCGCGCGCGCGCGCGCTGTCTTCGCGGCGCTGTGGTGGCC 167  
Db 774 GAGGTAGCGCAGCATTTCCCGCAGGCTCAGCGGCGCGCGCGCGCGCGCGCGCGCG 715  
QY 168 GTGGCGCTGTGCCAAGAGCGAGTTCAGCGCGCAGAGGCTCAGGGAACCTGAGTGTCTGGCAG 227  
Db 714 GCGCCCGCTCCGCGCGCGCTGTGGTTCGCGCGCGCGCGCGCTGTCTGTCTGTGGCGCGCG 655  
QY 228 GAGGTCCAGAGAGCGCGCTCGACGCGTCCGCCAGTCTCTGACCGCGGCGAGCTAACCGCG 287  
Db 654 CGCGCGCTGTGGCGCGCGCGCAGCGCGGCTCCCGGGCGCGACCGCGGGGGGGG 595  
QY 288 GCGCGCGCGCGCGCGCGCTTGGCCCGTTCCCGTGGGCGCACCGGGCTCCGGGATCGCGTGC 347  
Db 594 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 535  
QY 348 TGCCAGACGCTCCAGGACCTGAGCGCGAGTCCCGTGCCTGCGCGCGCGCGCGCGCGCGCG 407  
Db 534 GCGCAGGGCGCGCGCGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 475  
QY 408 AGGGGCTACGAGAGGCGCATGCGCGCGCTGGAGAGAGGCTGTGGTGGCCATGGGGCGCGCAG 467  
Db 474 CGCTGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 415  
QY 468 CAGCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 527  
Db 414 CGCGCGCGTGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 355  
QY 528 CGCGCAGGAGAGGAGTACCGGTACCGTACCGTACCGTACCGTACCGTACCGTACCGTACCG 562  
Db 354 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 320  
RESULT 6  
AAH26495/c  
ID AAH26495 standard; DNA; 12425 BP.  
XX  
AC AAH26495;  
XX  
DT 12-NOV-2001 (first entry)  
XX  
DE Human low density lipoprotein binding protein 2 (LBP-2) gene.  
XX  
KW Low density lipoprotein binding protein 2; LBP-2; LDL; human;  
KW atherosclerosis; antiarteriosclerotic; gene therapy; diagnosis;  
KW vaccine; ds.  
XX  
OS Homo sapiens.  
XX  
PH Key Location/Qualifiers  
FT CDS 2832..5153  
FT /tag= a  
FT /note= "includes introns"  
FT exon 2832..3785  
FT /tag= b  
FT intron 3786..4207  
FT /tag= c  
FT exon 4208..4502  
FT /tag= d  
FT intron 4503..4593  
FT /tag= e  
FT exon 4594..4694  
FT /tag= f  
FT intron 4695..4787  
FT /tag= g  
FT exon 4788..4899  
FT /tag= h  
FT intron 4900..4994  
FT /tag= i  
FT exon 4995..5153  
FT /tag= j  
XX WO200164874-A2.  
PN



QY 428 GCCGCGCTGAGAAAGCTGTGGCCATCGGGGCGGCGACAGCAGCGCGCGCGGCGAGG 487  
Db 2287 TCTGGAGAGCAGGTCCGCGCGCGCGCGCGGAGCTCAGAGCGCGGCTCGCG 2228  
QY 488 AGGAGGAGGAGGACAGGGGGCTACTACTACCCCTGTCAGCGCGCGCGGAGGAGATACGG 547  
Db 2227 GCAGCGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2168  
QY 548 CTACGGTCAGGTGGCGAGCGCGGAGATGATCACCCTGTGCTCCCGGCGACACCGCGG 607  
Db 2167 CTTCTGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2113  
QY 608 CGGCGCAAGGATCGCGCGCGGTGAGGCTTACGAAGGCGCGGAGTACGCGCGGGTTGCC 667  
Db 2112 GTCGCGCAGAACCGAGCTCGCGGAGCGCGCGCGCGCGCGCGCGCGCGCG 2053  
QY 668 GATGATGTCCCGCTGTGCGAGCCC 692  
Db 2052 CGGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2028

RESULT 8  
ID AAD25519 standard; DNA; 154746 BP.

AC AAD25519;  
XX  
DT 26-MAR-2002 (first entry)  
XX  
DE Human herpesvirus 2 complete DNA genome.  
DE  
KW Human herpesvirus 2; cytostatic; cancer; immunosuppressive; virucide;  
KW antibacterial; fungicide; protozoicide; antirheumatic; antiinflammatory;  
KW antiarthritic; rheumatoid arthritis; neuroprotective; multiple sclerosis;  
KW immune response; vasotropic; vaccine; gene therapy; autoimmune disease;  
KW vasculitis; ds.  
XX  
OS Human herpesvirus 2.  
XX  
PN WO200176643-A1.  
XX  
PD 18-OCT-2001.  
XX  
PE 06-APR-2001; 2001WO-US11372.  
XX  
PR 07-APR-2000; 2000US-195680P.  
XX  
PA (BAYU ) BAYLOR COLLEGE MEDICINE.  
XX  
PI Orson FM, Kinsey BM, Bhogal BS;  
XX  
DR WPI; 2002-066308/09.  
XX

Composition for oral delivery of vaccines, comprises expression vector containing antigenic genomic sequence, bound to aggregated protein-polycationic polymer conjugate or suspension - - -  
Disclosure; Page 90-132; 145pp; English.  
XX  
The invention relates to a composition comprising an expression vector bound to an aggregated protein-polycationic polymer conjugate or suspension. The expression vector contains a promoter polynucleotide sequence operatively linked to a polynucleotide sequence encoding an antigen which is a fragment of a gene or genome associated with an infectious disease, cancer and autoimmune disease such as rheumatoid arthritis, vasculitis, and multiple sclerosis, pathogenic genomes consisting of bacterium, fungus, protozoa and virus such as human immunodeficiency virus (HIV), herpes simplex virus (HSV), hepatitis C virus (HCV), influenza and respiratory syncytial virus (RSV), and optionally comprising a nucleotide sequence encoding a cytokine (or a cytokine expression vector), is useful for inducing an immune response (systemic and/or mucosal) in an organism. The cytokine expression vector

contains a sequence for granulocyte macrophage-colony stimulating factor (GM-CSF) or interleukin-12 (IL-12). The polynucleotide sequences encoding the antigen and the cytokine are under transcriptional control of same or different promoter polynucleotide sequences. The expression vector, as a DNA vaccine is useful for treating a condition in an organism. The present sequence is human herpesvirus 2 complete DNA genome related to the invention.

Sequence 154746 BP; 23003 A; 54218 C; 54701 G; 22824 T; 0 other;

Query Match 6.7%; Score 63.4; DB 24; Length 154746;  
Best Local Similarity 45.8%; Pred. No. 0.027;  
Matches 259; Conservative 0; Mismatches 301; Indels 5; Gaps 1;

QY 128 GCGCGCGCGCGCGCGCTGCTTCGCGCGCTGTGGCGCGCTGTGCGCGCTGTGCGCGCGA 187  
Db 129446 GCGCGCGCGCGCGCGCTGCGGCTCGGGCCCTTCGCGCGCGCGCGCGCGCGCGCG 129505  
QY 188 GGTTCGAGCGCGCAGAGGCTCAGGGACCTGCAGTGTCTGGCAGGAGGTCACGAGAGCCCGT 247  
Db 129506 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 129565  
QY 248 CGACGGTCCCGCAGGTCTTCGACCGGAGTAAACCGCGCGCGCGCGCGCGCGCGCG 307  
Db 129566 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 129625  
QY 308 TGGCCCGTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 367  
Db 129626 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 129685  
QY 368 GAGCGCGCGAGTCCGCTGCGCGCGCATCCGAGCATGTGTGCGCGCGCTACGAGGAGGCCAT 427  
Db 129686 GCGCGCGCGAGCGAGTCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 129745  
QY 428 GCGCGCGCTGGAGAAAGGTGTGTGCGCATGTGGCGCGCGCGCGCGCGCGCGCGCG 487  
Db 129746 TCTGGAAGAGCAGGTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 129805  
QY 488 AGGAGGAGGAGCAGGGGGGCTACTACTACCCCTGCGAGCGCGCGCGCGCGCGCGCG 547  
Db 129806 GCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 129865  
QY 548 CTACGCTCAGGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 607  
Db 129866 CTTCTGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 129920  
QY 608 CGGCGCAAGGATCGCGCGCGGTGAGGCTTACGAAGCGCGCGCGCGCGCGCGCGCG 667  
Db 129921 GTCGCGCAGAAACCGCAGCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 129980  
QY 668 GATGATGTCCCGCTGTGCGAGCCCC 692  
Db 129981 CGGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 130005

RESULT 9  
AAD25519/c  
ID AAD25519 standard; DNA; 154746 BP.

XX  
AC AAD25519;  
XX  
DT 26-MAR-2002 (first entry)  
XX  
DE Human herpesvirus 2 complete DNA genome. .  
XX  
KW Human herpesvirus 2; cytostatic; cancer; immunosuppressive; virucide;  
KW antibacterial; fungicide; protozoicide; antirheumatic; antiinflammatory;  
KW antiarthritic; rheumatoid arthritis; neuroprotective; multiple sclerosis;  
KW immune response; vasotropic; vaccine; gene therapy; autoimmune disease;  
KW vasculitis; ds.  
XX  
OS Human herpesvirus 2.  
XX





XX	AA09010 standard; cDNA; 1272 BP.	Matches 192; Conservative 0; Mismatches 196; Indels 2; Gaps 1;
AC	AA09010;	
DT	14-JUN-1999 (first entry)	
XX	Brn-3a polynucleotide.	
DE		
XX	Brn-3a; Bcl-2; neurons; neuronal cells; apoptosis; cell death; CNS;	
KW	PNS; central nervous system; parasympathetic nervous system;	
KW	development; injury; neurotrophic factor; nerve growth factor; NGF;	
KW	ciliary neurotrophic factor; CNRF; brain-derived neurotrophic factor;	
KW	BNTF; neurotrophin; NT-3; NT-4; NT-5; neurodegenerative disease;	
KW	familial dysautonomia; infantile muscular dystrophy;	
KW	Parkinson's disease; Alzheimer's disease; ss.	
XX		
OS	Homo sapiens.	
XX		
FH	Key	Location/Qualifiers
FT	CDS	1..1272
FT		/*tag= a
FT		/product= Brn_3a_polypeptide
XX		
XX	WO9905272-A1.	
PN		
PD	04-FEB-1999.	
XX		
XX	27-JUL-1998; 98WO-GB02228.	
PF		
XX		
PR	10-DEC-1997; 97US-0988476.	
PR	25-JUL-1997; 97GB-0015823.	
XX		
PA	(UNLO ) UNIV COLLEGE LONDON.	
XX		
PI	Latchman DS, Smith MD;	
XX		
DR	WPI; 1999-142928/12.	
DR	P-PSDB; AAW96262.	
XX		
XX	New polypeptide comprising transcription factor Brn-3a, or its	
PT	derivative - useful for treating nervous system diseases, preventing	
PT	cellular apoptosis and increasing nerve regeneration following	
PT	neuronal damage	
XX		
PS	Disclosure; Page 61-62; 68pp; English.	
XX		
CC	Over expression of transcription factor Brn-3a can protect neuronal	
CC	cells from apoptosis. Brn-3a also specifically activates expression	
CC	of the Bcl-2 gene in neuronal cells and this activation is mediated	
CC	via a Brn-3a response element in the 5' regulatory region of the	
CC	Bcl-2 gene. Both the anti-apoptotic effect of Brn-3a and its ability	
CC	to activate expression of Bcl-2 are mediated by the N-terminal	
CC	domain of Brn-3a. Members of the Bcl-2 family perform critical roles	
CC	in the regulation of selective apoptosis during development of the	
CC	nervous system. The stimulation of Bcl-2 expression by Brn-3a in a	
CC	neuron specific manner and consequent protection of neuronal cells	
CC	from apoptosis suggests that Brn-3a may co-ordinate some aspects of	
CC	neuronal reorganisation during development or following injury. The	
CC	elevation of Brn-3a expression by either pharmacological means	
CC	(compositions comprising one other therapeutic polypeptide e.g.	
CC	neurotrophic factors, nerve growth factor (NGF), ciliary neurotrophic	
CC	factor (CNTF), brain-derived neurotrophic factor (BDNF), and	
CC	neurotrophins NT-3 and NT-4/5) or gene therapy may represent a	
CC	method for treating human diseases associated with excessive	
CC	neuronal cell death and/or lack of nerve regeneration, especially	
CC	neurodegenerative diseases such as familial dysautonomia and	
CC	infantile muscular dystrophy, and Parkinson's and Alzheimer's	
CC	disease.	
XX		
SQ	Sequence 1272 BP; 202 A; 468 C; 452 G; 150 T; 0 other;	
XX		
XX	Query Match	6.6%; Score 62.4; DB 20; Length 1272;
XX	Best Local Similarity	49.2%; Pred. No. 0.03;

QY	142	CGCTGTGCTTCGGCGCCCTGGTGGCGCGTCTGCGCAAGCGGAGGTCGAGCGGCGAGA	201
DB	365	CGCGCTGCTCGGCTCATGGCGCGCGGGCGCGGGCGCGGGCGCGGGCGCGGGCG	424
QY	202	GGCTCAGGACCTGCAGTGTGGCAGGAGGTCCAGAGAGCCCGTCTCGACCGTGCCTGC	261
DB	425	GGGGCGGGCGGCCACGACGGCCCGGGGGCGGTGGCGCCCGCGCGCGCGCGCGCC	484
QY	262	AGTCTCTGACCGGAGTAAACGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGG	321
DB	485	CGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCG	544
QY	322	GGGGCACCGGGCTCCGGATGCGGTCTGCGACAGCTCCAGGACGTGAGCCCGCGAGTGC	381
DB	545	CGGGCGGGCGGGCTCTGGCGGGCTCGCGGCGCTCCGCGCACCTCACCGCATATGCA	604
QY	382	GCTGCGCGCCCATCCGAGCATGTTGTCAGGGGCTACGAGGAGCCATGCGCGCGCTGG	441
DB	605	ACCTGTGCGACCCCGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCG	662
QY	442	AAGGCTGTGTCATGGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCG	501
DB	663	CGGGCTGTGTCGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCG	722
QY	502	AGGGGGGCTACTACTACCCCTGCGAGCCGCGC	531
DB	723	CGGGCAGGTGGCAGCGGCGCATCGCGCGCGC	752
XX	RESULT 12		
XX	AAA29006		
ID	AAA29006 standard; cDNA; 1272 BP.		
XX	AAA29006;		
XX	12-SEP-2000 (first entry)		
XX	Human transcription factor Brn-3a coding sequence.		
XX	Brn-3a; modulator; inhibitor; cervical cancer; human papilloma virus;		
XX	HPV; antisense; cytostatic; ss.		
XX	Homo sapiens.		
XX			
FH	Key	Location/Qualifiers	
FT	CDS	1..1272	
FT		/*tag= a	
FT		/product= Brn-3a_transcription_factor	
XX			
PN	WO200034466-A1.		
XX			
PD	15-JUN-2000.		
XX			
PF	07-DEC-1999; 99WO-GB04116.		
XX			
PR	07-DEC-1998; 98GB-0026888.		
PR	31-MAR-1999; 99US-0282210.		
XX			
PA	(UNLO ) UNIV COLLEGE LONDON.		
XX			
PI	Latchman DS, Budhram-Mahadeo V, Ndisang D;		
XX			
DR	WPI; 2000-423416/36.		
DR	P-PSDB; AAY96404.		
XX			
PT	Product for treating, preventing and diagnosing cervical cancer		
PT	comprises a nucleotide sequence or molecule which binds to Brn-3a,		
PT	decreases its intracellular levels or inhibits its activity		
XX			
PS	Disclosure; Page 63-65; 72pp; English.		
XX			









GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 29, 2003, 14:01:53 ; Search time 75 Seconds  
(without alignments)  
5590.853 Million cell updates/sec

Title: US-10-053-410-3

Perfect score: 950

Sequence: 1 aaaaaaacccctcgtcat.....aaaaaaaaaaaaaaaaaaaaa 950

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 569978 seqs, 220691566 residues

Total number of hits satisfying chosen parameters: 1139956

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents NA:\*

- 1: /cgn2\_6/ptodata/2/ina/5A COMB.seq.\*
- 2: /cgn2\_6/ptodata/2/ina/5B COMB.seq.\*
- 3: /cgn2\_6/ptodata/2/ina/6A COMB.seq.\*
- 4: /cgn2\_6/ptodata/2/ina/6B COMB.seq.\*
- 5: /cgn2\_6/ptodata/2/ina/PCTUS COMB.seq.\*
- 6: /cgn2\_6/ptodata/2/ina/backfile1.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	91.2	9.6	2296	2	US-08-899-336-2
2	59.4	6.3	4524	2	US-08-845-998-7
3	59.4	6.3	4524	3	US-09-206-537-7
4	59.4	6.3	4524	3	US-09-430-854-7
5	59	6.2	3056	4	US-09-016-434-1153
6	57.4	6.0	3169	4	US-09-820-312B-713
7	57.4	6.0	4190	3	US-08-924-345-1
8	56.4	5.9	1556	3	US-09-043-937A-3
9	55.4	5.8	4403765	3	US-09-103-840A-2
10	55	5.8	4411529	3	US-09-103-840A-1
11	53.4	5.6	530	3	US-08-758-662-4
12	52.8	5.6	672	4	US-09-352-991A-14411
13	52.8	5.6	888	4	US-09-352-991A-14356
14	52.8	5.6	894	4	US-09-352-991A-6276
15	52.8	5.6	1530	4	US-09-352-991A-14293
16	52.8	5.6	3594	4	US-09-252-991A-14262
17	52.6	5.5	8438	1	US-07-945-283-1
18	52.4	5.5	1401	4	US-09-352-991A-14814
19	52.4	5.5	1857	4	US-09-352-991A-7125
20	52.4	5.5	1908	4	US-09-352-991A-7125
21	52.4	5.5	2925	4	US-09-352-991A-7069
22	52.4	5.5	3900	4	US-09-352-991A-14550
23	52.2	5.5	152331	3	US-09-128-155-16
24	52.2	5.5	4403765	3	US-09-103-840A-2
25	52.2	5.5	4411529	3	US-09-103-840A-1
26	51	5.4	525	4	US-09-252-991A-15328
27	51	5.4	714	4	US-09-252-991A-1104

28	51	5.4	759	4	US-09-252-991A-982	Sequence 982, Appl
29	51	5.4	957	4	US-09-252-991A-15376	Sequence 15376, A
30	51	5.4	963	4	US-09-252-991A-15461	Sequence 15461, A
31	51	5.4	1119	4	US-09-252-991A-3537	Sequence 3537, Ap
32	51	5.4	1289	2	US-08-344-833-1	Sequence 1, Appli
33	51	5.4	1467	4	US-09-252-991A-1142	Sequence 1142, Ap
34	51	5.4	1581	4	US-09-252-991A-3549	Sequence 3549, Ap
35	51	5.4	1947	4	US-09-252-991A-3557	Sequence 3557, Ap
36	50.8	5.3	966	4	US-09-252-991A-14511	Sequence 14511, A
37	50.8	5.3	969	4	US-09-252-991A-14770	Sequence 14770, A
38	50	5.3	1863	4	US-09-252-991A-3225	Sequence 3225, Ap
39	50	5.3	1826	4	US-09-249-585A-2	Sequence 2, Appli
40	50	5.3	1926	4	US-09-410-399-3	Sequence 3, Appli
41	50	5.3	2154	4	US-09-252-991A-3346	Sequence 3346, Ap
42	50	5.3	2580	3	US-09-050-863-2	Sequence 2, Appli
43	50	5.3	2580	4	US-09-359-081-2	Sequence 2, Appli
44	50	5.3	5452	2	US-09-130-114-1	Sequence 1, Appli
45	50	5.3	8705	4	US-09-647-344A-14	Sequence 14, Appli

#### ALIGNMENTS

#### RESULT 1

US-08-899-336-2  
; Sequence 2, Application US/08899336

; Patent No. 5955649

; GENERAL INFORMATION:

; APPLICANT: HIROTA, NAHIKO

; APPLICANT: KIHARA, MAKOTO

; APPLICANT: KURODA, HISAO

; APPLICANT: ITO, KAZUTOSHI

; TITLE OF INVENTION: GENE EXPRESSION REGULATORY DNA,

; TITLE OF INVENTION: EXPRESSION CASSETTE, EXPRESSION VECTOR AND TRANSGENIC

; TITLE OF INVENTION: PLANT

; NUMBER OF SEQUENCES: 8

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: P.C.

; ADDRESS: 1755 SOUTH JEFFERSON DAVIS HIGHWAY, SUITE 400

; CITY: ARLINGTON

; STATE: VIRGINIA

; COUNTRY: U.S.A.

; ZIP: 22202

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patent In Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/899,336

; FILING DATE: 23-JUL-1997

; CLASSIFICATION: 800

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: JP HEI 8-193433

; FILING DATE: 23-JUL-1996

; ATTORNEY/AGENT INFORMATION:

; NAME: OBLON, NORMAN F.

; REGISTRATION NUMBER: 24,618

; REFERENCE/DOCKET NUMBER: 2589-0061-0

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (703) 413-3000

; TELEFAX: (703) 413-2220

; INFORMATION FOR SEQ ID NO: 2:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 2296 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: double

; TOPOLOGY: linear

; MOLECULE TYPE: DNA (genomic)

; US-08-899-336-2

Query Match 9.6%; Score 91.2; DB 2; Length 2296;



```

; LENGTH: 4524 base pairs
;
; TYPE: nucleic acid
;
; STRANDEDNESS: double
;
; TOPOLOGY: linear
;
; MOLECULE TYPE: CDNA
;
; HYPOTHETICAL: NO
;
; ANTI-SENSE: NO
;
; FEATURE:
;
; NAME/KEY: CDS
;
; LOCATION: 174..1433
;
US-09-206-537-7

```

Query Match 6.3%; Score 59.4; DB 3; Length 4524;  
Best Local Similarity 48.8%; Pred. NO. 0.0072;  
Matches 189; Conservative 0; Mismatches 196; Indels 2; Gaps 1;

RESULT 4  
US-09-430-854-7  
; Sequence 7, Application US/09430854  
; Patent No. 6271019  
; GENERAL INFORMATION:  
; APPLICANT: Van Baren, Nicolas  
; APPLICANT: Coulie, Pierre G.  
; APPLICANT: De Smet, Charles  
; APPLICANT: Lucas, Sophie  
; APPLICANT: Boon, Thierry  
; TITLE OF INVENTION: LEUKEMIA ASSOCIATED GENES  
; NUMBER OF SEQUENCES: 16  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Wolf, Greenfield & Sacks, P.C.  
; STREET: 600 Atlantic Avenue  
; CITY: Boston  
; STATE: MA  
; COUNTRY: US  
; ZIP: 02210  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/430,854  
; FILING DATE:  
; CLASSIFICATION:

PRIOR APPLICATION DATA: 08/845,998  
 APPLICATION NUMBER: 08/845,998  
 FILING DATE:  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Van Amsterdam, John R.  
 REGISTRATION NUMBER: 40,212  
 REFERENCE/DOCKET NUMBER: L0461/7008  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (617) 720-3500  
 TELEFAX: (617) 720-2441  
 INFORMATION FOR SEQ ID NO: 7:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 4524 base pairs  
 TYPE: nucleic acid  
 STRANDEDNESS: double  
 TOPOLOGY: linear  
 MOLECULE TYPE: CDNA  
 HYPOTHETICAL: NO  
 ANTI-SENSE: NO  
 FEATURE:  
 NAME/KEY: CDS  
 LOCATION: 174..1433  
 US-09-430-854-7

Query Match	6.3%;	Score 59.4;	DB 3;	Length 4524;
Best Local Similarity	48.8%;	Pred. No. 0.0072;		
Matches 189; Conservative	0;	Mismatches 196;	Indels 2;	Gaps 1
<hr/>				
Qy	145	TGTGCTTCGGCGCCCTGGTGGCGCTGTGCCTCTGCCAAGACGAGGTTCGAGCGCAGAGGC	204	
Db	532	TCTCTTCGCGCTTGCTTCGCGCTATATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG	591	
<hr/>				
Qy	205	TCAGGGAACCTGCAGTGTCTGTCAGAGAGTCCAGGAGAGCCCCGTTCGACGCGFTGCCGCCACGG	264	
Db	592	GCGCGCGCGGCCACGACGCGCCGCGGGCGCGGTGCGCGCCC CGCGCGCGCGCGCGCGCGCGCG	651	
<hr/>				
Qy	265	TCCTCGAACCGGACGTAAACCGCGCGCGCGCGCGCGCGCGCGTTGGGCCCGTTCCCGGTGGG	324	
Db	652	GCGCGCGCGCGCCCCCGGGAGGCGCGCGGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG	711	
<hr/>				
Qy	325	GCACCGGGCTCCCGATGCGGTGTGTCACAGCAGCTCCAGAGACGTGAGCGCGCAGTGCCTCCT	384	
Db	712	GCGCGGGCTCCTTGGCGGTTCGCGCACCTCACCCGCATATGCACAGCTTGGGCCACC	771	
<hr/>				
Qy	385	GCGCGCGCATTCGGAAGCATGGTCAGGGGCTACGAGAGAGGCCATGCCCGCGCTGGAGAAG	444	
Db	772	TGTGCAACCCCGCGCGCGCGGCCCATGAACATGCCGTTCGGGCTGCCGC--ACCCTGG	829	
<hr/>				
Qy	445	GCTGTTGGCCATCGGGGCGCGACAGCAGCGCGCGCGCGAGGAGGAGGAGGACAGG	504	
Db	830	GCTGTTGGCGCGCGCGCGCACACAGCGCGCGCACAGCGCGCGCGCGCGCGCGCGCGCGCG	889	
<hr/>				
Qy	505	GGGGCTACTACTACCCCTGCAGCGCG	531	
<hr/>				
Db	890	GCAAGTTGGCAGCGGCATCGCGCGCGCG	916	

RESULT 5  
US-09-016-434-1153/c  
; Sequence 1153, Application US/09016434  
; Patent No. 6500938  
; GENERAL INFORMATION:  
; APPLICANT: Janice Au-Young  
; APPLICANT: Jeffrey J. Seilhamer  
; TITLE OF INVENTION: COMPOSITION FOR THE DETECTION OF SIGNALING  
; TITLE OF INVENTION: PATHWAY GENE EXPRESSION  
; NUMBER OF SEQUENCES: 1490  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: INCYTE PHARMACEUTICALS, INC.  
; STREET: 3174 PORTER DRIVE  
; CITY: PALO ALTO  
; STATE: CALIFORNIA  
; COUNTRY: USA







Query Match 5.9%; Score 56.4; DB 3; Length 1556;  
Best Local Similarity 48.4%; Pred. No. 0.024;  
Matches 156; Conservative 0; Mismatches 166; Indels 0; Gaps 0;

QY 101 CTGAGAGAGATGGCTAAGATCCCGCGGCGCGCGCGCGCGCGCTGTGCTTCCGCGCCCT 160  
DB 740 CGGCAAGGCGGTGGCGAGCAGCGCGCTGGAGGCGCGCGCGCGCGCGCTGGACCT 799

QY 161 GGTGGCGGTGGCGCTGTGCAAGGCGAGGTGAGCGGAGAGCTCAGGACCTGCAATG 220  
DB 800 GGTGGGTGGTGAACCCCTGTGTGTGGCGCCCTGCTGTCAGGCGACAGGTGAACCCAG 859

QY 221 CTGGCAGGAGGTCCAGAGAGCCGCTCGACGCGTGGCGCGGCTCTCGACCGGAGCT 280  
DB 860 CATCGCGCATCTCTAAGTACCTGACGCTGGCGCGGACCTTTCGCCAAGCCCGTGA 919

QY 281 AACCGCGCGCGCGCGCGCGCGCTGGCGCGTTCGCGTGGGCGACCGGCTCCGGAT 340  
DB 920 GCGGTACGTGGAGCTGGCGGACGTGGCGGACGCGGACCTCCGCGTCTTCGAGAGCCCG 979

QY 341 GCGGTGCTGCCACAGCTCCAGAGCTGAGCGCGGAGTGGCGTGGCGCGGCTCCGGAG 400  
DB 980 CGGTCGCGCGCGCACCTCTGCGCGGAGCGCTCTCCACCGCGGAGAGCTGCTCGGCAT 1039

QY 401 CATGGTCAGGGGTACGAGGAG 422  
DB 1040 CCTCGCAAGCTTCTCCCGAG 1061

RESULT 9  
US-09-103-840A-2  
; Sequence 2, Application US/09103840A  
; Patent No. 6294328  
; GENERAL INFORMATION:  
; APPLICANT: FLEISCHMAN, Robert D.  
; APPLICANT: WHITE, Owen R.  
; APPLICANT: FRASER, Claire M.  
; APPLICANT: VENTER, John C.  
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM  
; FILE REFERENCE: 24366-20007.00  
; CURRENT APPLICATION NUMBER: US/09/103,840A  
; CURRENT FILING DATE: 1998-06-24  
; NUMBER OF SEQ ID NOS: 2  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 2  
; LENGTH: 4403765  
; TYPE: DNA  
; ORGANISM: Mycobacterium tuberculosis  
; FEATURE:  
; OTHER INFORMATION: CDC 1551  
; OTHER INFORMATION: "n" bases at various positions throughout the sequence  
; OTHER INFORMATION: represent a, t, c or g

US-09-103-840A-2

Query Match 5.8%; Score 55.4; DB 3; Length 4403765;  
Best Local Similarity 45.9%; Pred. No. 0.076;  
Matches 230; Conservative 0; Mismatches 266; Indels 5; Gaps 1;

QY 69 CGACAGCAAAAGATTTGTGAGGCTGATCCCGCTGAGAGAGATGGCTAAGATCGCGCG 128  
DB 3941279 CCACAGGCAACGCGGTGTGCGCGGCGCGCGCGCGCGCGGAGCGCGCG 3941338

QY 129 GCGCGCGCGCGCGCTGTGCTTTGCGCGCCCTGGTGGCGGCTCGCTGCTGCTGCTGCTG 188  
DB 3941339 GGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3941398

QY 189 GTGAGGCGCGAGGCTCAGGACCTGAGTGTGCGAGAGGTCCAGAGAGCCGCTC 248  
DB 3941399 -----CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3941453

QY 249 GAGCGGTGCGCGCGAGTCTCTGACCGCGAGCTAACCGCGCGCGCGCGCGCGCGCG 308

DB 3941454 CAACGCGCTCCGCGCGCGCGCGCGCACTGCGGACAAAGCGCGCGCGGTGCTGGCG 3941513

QY 309 GGCCTGTTCCGCTGGGCGACCGGCTCCGGATCGGTGCTGCCAGCAGCTCCAGGACGTG 368

DB 3941514 GGCCTGCGCGGATACCCACCGGATCGCGCGGCGCGCGCGGTGACCGCGCGCACCGGCG 3941573

QY 369 AGCCGCGAGTGGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 428

DB 3941574 AGCGCGCGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3941633

QY 429 CGCGCGCTGGAGAAAGTGTGGCGATGGGCGCGCGCGCGCGCGCGCGCGCGCGCG 488

DB 3941634 CGGCACCAAGCGCGGTGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3941693

QY 489 GGAGGAGGAGGAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 548

DB 3941694 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3941753

QY 549 TACGCTCAGGGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 569

DB 3941754 TCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3941774

RESULT 10  
US-09-103-840A-1  
; Sequence 1, Application US/09103840A  
; Patent No. 6294328  
; GENERAL INFORMATION:  
; APPLICANT: FLEISCHMAN, Robert D.  
; APPLICANT: WHITE, Owen R.  
; APPLICANT: FRASER, Claire M.  
; APPLICANT: VENTER, John C.  
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM  
; FILE REFERENCE: 24366-20007.00  
; CURRENT APPLICATION NUMBER: US/09/103,840A  
; CURRENT FILING DATE: 1998-06-24  
; NUMBER OF SEQ ID NOS: 2  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 1  
; LENGTH: 4411529  
; TYPE: DNA  
; ORGANISM: Mycobacterium tuberculosis  
; OTHER INFORMATION: H37Rv

US-09-103-840A-1

Query Match 5.8%; Score 55; DB 3; Length 4411529;  
Best Local Similarity 47.9%; Pred. No. 0.09;  
Matches 218; Conservative 0; Mismatches 235; Indels 2; Gaps 2;

QY 113 GGCTAAGATCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 172

DB 3947296 GGCCTAAGCGGTGACGCGCGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3947354

QY 173 CGTCTGCCAAGCGCGAGTTCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 232

DB 3947355 CACCGCGCGCACCGCGGTTCG-CGCGTGGAGCGCGCGCGCGCGCGCGCGCGCG 3947413

QY 233 CCAGGAGAGCGCGCTCGACCGCGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 292

DB 3947414 GCGGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3947473

QY 293 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 352

DB 3947474 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3947533

QY 353 GCAGTCCAGGAGCGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 412

DB 3947534 AGCGCGCGCACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3947593

QY 413 CTACGAGGAGCGCATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 472

DB 3947594 GCACCGCGCGCATGATCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3947653





GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 29, 2003, 15:12:39 ; Search time 353 Seconds  
(without alignments)  
8866.730 Million cell updates/sec

Title: US-10-053-410-3  
Perfect score: 950  
Sequence: 1 aaaaaaacccctcgtcgat.....aaaaaaaaaaaaaaaaaaaaa 950

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 2190069 seqs, 1647345023 residues

Total number of hits satisfying chosen parameters: 4380138

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

- Database : Published Applications NA.\*
- 1: /cgn2\_6/ptodata/2/pubpna/US07\_PUBCOMB.seq.\*
  - 2: /cgn2\_6/ptodata/2/pubpna/PCT\_NEW\_PUB.seq.\*
  - 3: /cgn2\_6/ptodata/2/pubpna/US06\_NEW\_PUB.seq.\*
  - 4: /cgn2\_6/ptodata/2/pubpna/US06\_PUBCOMB.seq.\*
  - 5: /cgn2\_6/ptodata/2/pubpna/US07\_NEW\_PUB.seq.\*
  - 6: /cgn2\_6/ptodata/2/pubpna/PCTUS\_PUBCOMB.seq.\*
  - 7: /cgn2\_6/ptodata/2/pubpna/US08\_NEW\_PUB.seq.\*
  - 8: /cgn2\_6/ptodata/2/pubpna/US08\_PUBCOMB.seq.\*
  - 9: /cgn2\_6/ptodata/2/pubpna/US09A\_PUBCOMB.seq.\*
  - 10: /cgn2\_6/ptodata/2/pubpna/US09B\_PUBCOMB.seq.\*
  - 11: /cgn2\_6/ptodata/2/pubpna/US09C\_PUBCOMB.seq.\*
  - 12: /cgn2\_6/ptodata/2/pubpna/US09\_NEW\_PUB.seq.\*
  - 13: /cgn2\_6/ptodata/2/pubpna/US10A\_PUBCOMB.seq.\*
  - 14: /cgn2\_6/ptodata/2/pubpna/US10B\_PUBCOMB.seq.\*
  - 15: /cgn2\_6/ptodata/2/pubpna/US10\_NEW\_PUB.seq.\*
  - 16: /cgn2\_6/ptodata/2/pubpna/US60\_NEW\_PUB.seq.\*
  - 17: /cgn2\_6/ptodata/2/pubpna/US60\_PUBCOMB.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	64.6	6.8	1614	10	US-09-976-740-45
C 2	64.6	6.8	1614	13	US-10-023-529-45
C 3	64.6	6.8	1614	13	US-10-023-523-45
C 4	64.6	6.8	12425	13	US-09-976-740-50
C 5	64.6	6.8	12425	13	US-10-023-529-50
C 6	64.6	6.8	12425	13	US-10-023-523-50
C 7	63.4	6.7	3957	12	US-10-200-562-193
C 8	63.4	6.7	3957	12	US-10-237-551-193
C 9	63.4	6.7	154746	12	US-09-827-688-8
C 10	63.4	6.7	154746	12	US-09-827-688-8
C 11	63	6.6	88421	10	US-09-976-059-1
C 12	61.2	6.4	3133	12	US-10-017-161-1483
C 13	59.4	6.3	3824	12	US-10-116-275-197
C 14	59	6.2	3035	10	US-09-864-864-311
C 15	59	6.2	3393	12	US-09-814-353-19995
C 16	57.4	6.0	3169	14	US-10-037-270-713

17	57.4	6.0	5452	12	US-10-017-161-1481	Sequence 1481, Ap
C 18	56.2	5.9	18435	14	US-10-156-761-412	Sequence 412, App
C 19	56.2	5.9	100000	14	US-10-156-761-15103	Sequence 15103, A
C 20	56.2	5.9	9025608	14	US-10-156-761-1	Sequence 1, Appli
21	56	5.9	2549	12	US-10-287-218-42	Sequence 42, Appl
22	56	5.9	2510	12	US-10-287-218-41	Sequence 41, Appl
23	55.6	5.9	15738	12	US-10-329-079-46	Sequence 46, Appl
C 24	55.6	5.9	61944	12	US-10-329-079-34	Sequence 34, Appl
C 25	55.4	5.8	2561	10	US-09-976-740-48	Sequence 48, Appl
C 26	55.4	5.8	2561	13	US-10-023-529-48	Sequence 48, Appl
C 27	55.4	5.8	2561	13	US-10-023-523-48	Sequence 48, Appl
28	55.4	5.8	18876	12	US-10-329-079-42	Sequence 42, Appl
29	54.6	5.7	3163	12	US-10-017-161-1857	Sequence 1857, Ap
C 30	54.4	5.7	1758	14	US-10-156-761-5340	Sequence 5340, Ap
C 31	53.8	5.7	1559	12	US-10-361-460-34	Sequence 34, Appl
C 32	53.8	5.7	3957	12	US-10-200-562-193	Sequence 193, App
C 33	53.8	5.7	3957	12	US-10-237-551-193	Sequence 193, App
34	53.6	5.6	873	12	US-10-029-386-20490	Sequence 20490, A
35	53.6	5.6	7155	12	US-10-329-079-14	Sequence 14, Appl
36	53.6	5.6	37360	12	US-10-329-079-6	Sequence 6, Appli
C 37	53.4	5.6	530	13	US-10-073-353-4	Sequence 4, Appli
C 38	53.4	5.6	42999	10	US-09-799-462A-17	Sequence 17, Appl
C 39	53.4	5.6	42999	11	US-08-836-911A-17	Sequence 17, Appl
C 40	53.4	5.6	42999	12	US-09-738-630-73	Sequence 73, Appl
C 41	53.4	5.6	42999	13	US-10-125-767-17	Sequence 17, Appl
C 42	53.4	5.6	42999	14	US-10-151-081-17	Sequence 17, Appl
C 43	53.4	5.6	42999	14	US-10-287-313-17	Sequence 17, Appl
C 44	53.4	5.6	42999	14	US-10-219-694-17	Sequence 17, Appl
C 45	53	5.6	1257	14	US-10-156-761-5999	Sequence 5999, Ap

ALIGNMENTS

RESULT 1

US-09-976-740-45/c  
; Sequence 45, Application US/09976740  
; Publication No. US20020194633A1  
; GENERAL INFORMATION:  
; APPLICANT: Lees, Ann M.  
; APPLICANT: Lees, Robert S.  
; APPLICANT: Law, Simon W.  
; APPLICANT: Arjona, Anibal A.  
; TITLE OF INVENTION: NOVEL LOW DENSITY LIPOPROTEIN BINDING  
; TITLE OF INVENTION: PROTEINS AND THEIR USE IN DIAGNOSING AND TREATING  
; TITLE OF INVENTION: ATHEROSCLEROSIS  
; FILE REFERENCE: 10797-004001  
; CURRENT APPLICATION NUMBER: US/09/976,740  
; CURRENT FILING DATE: 2001-10-12  
; PRIOR APPLICATION NUMBER: 09/616,289  
; PRIOR FILING DATE: 2000-07-14  
; PRIOR APPLICATION NUMBER: US 08/979,608  
; PRIOR FILING DATE: 1997-11-26  
; PRIOR APPLICATION NUMBER: US 60/031,930  
; PRIOR FILING DATE: 1996-11-27  
; PRIOR APPLICATION NUMBER: US 60/048,547  
; PRIOR FILING DATE: 1997-06-03  
; NUMBER OF SEQ ID NOS: 53  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 45:  
; LENGTH: 1614  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (1) ... (1614)  
US-09-976-740-45

Query Match 6.8%; Score 64.6; DB 10; Length 1614;  
Best Local Similarity 46.4%; Pred. No. 7.2e-06;  
Matches 211; Conservative 0; Mismatches 244; Indels 0; Gaps 0;

QY 108 GAGATGGCTAAGATCCCGCGGGCGGGCGGCTGTGTCGCGGCTGTGGCC 167







[illegible]

```

RESULT 6
US-10-023-523-50/c
; Sequence 50, Application US/10023523
; Publication No. US20020152485A1
; GENERAL INFORMATION:
; APPLICANT: Lees, Ann M.
; APPLICANT: Lees, Robert S.
; APPLICANT: Law, Simon W.
; APPLICANT: Arjona, Arribal A.
; TITLE OF INVENTION: NOVEL LOW DENSITY LIPOPROTEIN BINDING
; TITLE OF INVENTION: PROTEINS AND THEIR USE IN DIAGNOSING AND TREATING
; TITLE OF INVENTION: ATHEROSCLEROSIS
; FILE REFERENCE: 10797-004001
; CURRENT APPLICATION NUMBER: US/10/023,523
; CURRENT FILING DATE: 2001-12-17
; PRIOR APPLICATION NUMBER: US/09/616,289
; PRIOR FILING DATE: 2000-07-14
; PRIOR APPLICATION NUMBER: US 09/517,849
; PRIOR FILING DATE: 2000-03-02
; PRIOR APPLICATION NUMBER: US 08/979,608
; PRIOR FILING DATE: 1997-11-26
; PRIOR APPLICATION NUMBER: US 60/031,930
; PRIOR FILING DATE: 1996-11-27
; PRIOR APPLICATION NUMBER: US 60/048,547
; PRIOR FILING DATE: 1997-06-03
; NUMBER OF SEQ ID NOS: 53
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 50
; LENGTH: 12425
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-023-523-50

```

Qy	288	G G C G G C G G C G G C G G C G G T T C C G T T G G G G A C A C C G G G C T C C G G A T G C G G T G C	347
Db	3425	G C G C G C G G C C C G C C G G G G G C C A C C G C C G G G G A G C G G C T G G G C G C G G C G C G C G C	3366
Qy	348	T G C A C A G C T C C A G A C G T G A G C C G A G T G C C G C T G C G C C G C C A T C C G G A G A C A T G G T C	407
Db	3365	G G C C A G G G C G C G G C C C G C T G C G C C G G G G G C C C G C T G C G C C G G G C C A G C G A G G G G G	3306
Qy	408	A G G G G C T A C A G A G A G C C A T G C C C G C G T G G A G A A A G G C T G T G G C A T C G G G G G C G G C A G	467
Db	3305	C G T T G T G C G C G G C G G C G C C G C G G G G C C C G G G C G G G C G G C G G C G C G A C G G	3246
Qy	468	C A G C A G C C C C C G C A G A G A G A G A G A G A G A G A G G G G G C T A C T A C C C C T G C A G C	527
Db	3245	C G C G G G C G T G C G G C G G G C G G C G T G G C G C G C G C G G C G G C G G C G G C G G C G G C G G	3186
Qy	528	C G G C C A G A G A G A T A C G G C T A C G T C A G G T G G	562
Db	3185	G G C C C C G C G G G G C G C G C G C G G G G C G G G G	3151

```

RESULT 7
US-10-200-562-193/c
; Sequence 193, Application US/10200562
; Publication No. US20030165819A1
; GENERAL INFORMATION:
; APPLICANT: McGowen, Patrick
; APPLICANT: Hosken, Nancy A.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE DIAGNOSIS AND
; TITLE OF INVENTION: TREATMENT OF HERPES SIMPLEX VIRUS INFECTION
; FILE REFERENCE: 210121-538C2
; CURRENT APPLICATION NUMBER: US/10/200,562
; CURRENT FILING DATE: 2002-07-19
; NUMBER OF SEQ ID NOS: 212
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 193
; LENGTH: 3957
; TYPE: DNA
; ORGANISM: HSV2
US-10-200-562-193

```

Query Match	6.7%	Score 63.4	DB 12	Length 3957
Best Local Similarity	45.8%	Pred. No. 1.8e-05		
Matches 259	Conservative 0	Mismatches 301	Indels 5	Gaps 1
Qy	128	GGCGGCGGCGGCGCGCTGTGCTTCGCGGCGCCCTGTGTGGCGCGTGTGGCGCGCTGTGCCAAGCGCGA	187	
Db	2587	GGCGGCGCAGCCGCTTGGCGGTTCGGGCGCCTCGCGGGCGCGCGGGTCAGCGCCGCGG	2528	
Qy	188	GGTGAGCGCGCAGAGGCTCAGGACCTGTGCAAGTCTCTGGCAGGAGGTTCCAGGAGAGCCCGCT	247	
Db	2527	GGCGCGCGCGCCCGCGCGCGCGTTCGGCGGGCGCGGGCGCGCGCCCCCGCGCGGAGGG	2468	
Qy	248	CGACGGTGTCGCGCAGGTCTCTGACCGGACGTAAACCGCGCGGCGGCGCGCGCGCGCGCT	307	
Db	2467	CGCGCGGCGGCGGGGGCGTTCGCGCGGCTCTTCTTGGGGGGCGCGGGGCGCGCGCCG	2408	
Qy	308	TGCGCCGCTTCGGTGGGGGACCGGGTTCGGATGCGGTGTGCCAGAGCTCCAGAGCGT	367	
Db	2407	CGCGCGCCCTGGCGGCGGGGCGCTCTTGCGCTTGGCGGCTCCCGCGGCGCGAGCGG	2348	
Qy	368	GAGCGCGAGTCCGCTTGCGCGCCCATCCGGAGCATGCTCAGGGGCTACGAGGAGGCCAT	427	
Db	2347	GC CGCGGAGCGAGTCCGCGCGCGGCACGGTGTTCGCCACGACAGGGGGCGCAGGCTCTGGT	2288	
Qy	428	CGCGCCGCTGGAGAAAGGCTGTGTGGCCATGTGGGGGCGCAGCAGCGCGCGCGCGAGGG	487	
Db	2287	TCTGGAAGACAGGTCCGCGGCGCGCGCGCGCGGAGCTCACAGCGCGGGCTCCGCG	2228	
Qy	488	AGGAGGAGGAGGACAGGGGGGCTACTTACCTCTGACGCCCGCGCAGGAGGGATACGG	547	
Db	2227	GCACGCGCGGGCCCGGGCCCGCGACACAGGCTCACGGCGCGCACGCGGGCCACGCGCG	2168	

QY 548 CTACGTGTCAGGTCGCGGACGAGATGTATCAACCTGTCTCCGCGCACCAACCGCGG 607  
Db 2167 CTTGCTGTCGCGCGGACGCGAGGT-----CCCCGCGAGCGCATCAGCACCGAGCG 2113  
QY 608 CGGGCCCAAGATCGCGCGGTGAGGCTTACGAAGGCGCGGAGTACGCCCGGGTTGCC 667  
Db 2112 GTCCGCGACGAACCGAGCTCGCGACGACGCGGCGCGGCGGTTCGCGGTGCGG 2053  
QY 668 GATGATGTGCGCGCTGTGCGAGGCC 692  
Db 2052 CGGGCGCGCGCGCGCGCGCGCGCC 2028

RESULT 8  
US-10-237-551-193/c  
; Sequence 193, Application US/10237551  
; Publication No. US20030165820A1  
; GENERAL INFORMATION:  
; APPLICANT: Day, Craig H.  
; APPLICANT: Hosken, Nancy A.  
; APPLICANT: Parsons, Joseph M.  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE DIAGNOSIS AND  
; TITLE OF INVENTION: TREATMENT OF HERPES SIMPLEX VIRUS INFECTION  
; FILE REFERENCE: 210121.538C3  
; CURRENT APPLICATION NUMBER: US/10/237,551  
; CURRENT FILING DATE: 2002-09-06  
; NUMBER OF SEQ ID NOS: 254  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 193  
; LENGTH: 3957  
; TYPE: DNA  
; ORGANISM: HSV2  
US-10-237-551-193

Query Match 6.7%; Score 63.4; DB 12; Length 3957;  
Best Local Similarity 45.8%; Pred. No. 1.8e-05;  
Matches 259; Conservative 0; Mismatches 301; Indels 5; Gaps 1;  
QY 128 GCGCGCGCGCGCGCGCTGTGCTTCGCGGCTGCTGCGCGCTGCGCGCTGCGCGCGCA 187  
Db 2587 GCGCGCGCGCGCGCGCTGCGCGCTGCGCGCTGCGCGCTGCGCGCTGCGCGCGCG 2528  
QY 188 GGTCTGAGCGCGAGAGGCTCAGGACCTGCGAGTCTGCGAGAGGCTCAGGAGAGCGCT 247  
Db 2527 GCGCGCGCGCGCGCGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2468  
QY 248 CGAGCGTTCGCGCGAGTCTCAGCGCGAGTAAACGCGCGCGCGCGCGCGCGCGCG 307  
Db 2467 CGCGCGCGCGCGCGCGCGCTGCGCGCGCTCTTCTTCTGCGCGCGCGCGCGCGCG 2408  
QY 308 TGGCCCGTTCGCGTGGGCGACCGCGCTCCGAGTCCGCGTGTGCGAGCGCTCCAGGACGT 367  
Db 2407 GCGCGCGCTGCGCGCGCGCGCGCTCTTCTGCGCTTGGCGCTTCCGCGCGCGAGCGG 2348  
QY 368 GAGCGCGAGTTCGCGTCCGCGCGCGCTCAGGAGTGTGCGAGCGCTCAGGAGGCGCAT 427  
Db 2347 GCGCGCGAGAGTTCGCGCGCGCGCGCGCTGCGCGCGCGCGCGCGCGCGCGCGCG 2288  
QY 428 GCGCGCGTGGAGAAAGTGTGTGCGCGTTCGCGCGCGCGCGCGCGCGCGCGCGCG 487  
Db 2287 TCTGGAAGAGCAGTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2228  
QY 488 AGGAGGAGGAGGAGCGGGGCTTACTACTACCTCCGCGCGCGCGCGCGCGCGCGCG 547  
Db 2227 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2168  
QY 548 CTACGCTGAGGTCGCGCGCGAGATGTATCAACCTGTCTGCGCGCGCGCGCGCGCG 607  
Db 2167 CTTGCTGTCGCGCGCGCGCGCGAGGT-----CCCCGCGAGCGCATCAGCACCGAGCG 2113  
QY 608 CGGGCCCAAGATCGCGCGGTGAGGCTTACGAAGGCGCGGAGTACGCCCGGGTTGCC 667  
Db 2112 GTCCGCGACGAACCGAGCTCGCGACGCGCGCGCGCGCGCGCGCGCGCGCGCG 2053

QY 668 GATGATGTGCGCGCTGTGCGAGGCC 692  
Db 2052 CGGGCGCGCGCGCGCGCGCGCGCC 2028

RESULT 9  
US-09-827-688-8  
; Sequence 8, Application US/09827688  
; Publication No. US20030165476A1  
; GENERAL INFORMATION:  
; APPLICANT: ORSON, FRANK  
; APPLICANT: KINSEY, BERNA  
; APPLICANT: BHOOGAL, BALBIR  
; TITLE OF INVENTION: MACROAGGREGATED PROTEIN CONJUGATES AS ORAL GENETIC IMMUNIZATION DI  
; TITLE OF INVENTION: AGENTS  
; FILE REFERENCE: P01949US1/10004014  
; CURRENT APPLICATION NUMBER: US/09/827,688  
; CURRENT FILING DATE: 2001-04-06  
; PRIOR APPLICATION NUMBER: 60/195,680  
; PRIOR FILING DATE: 2000-04-07  
; NUMBER OF SEQ ID NOS: 13  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 8  
; LENGTH: 154746  
; TYPE: DNA  
; ORGANISM: HERPESVIRUS 2  
US-09-827-688-8

Query Match 6.7%; Score 63.4; DB 12; Length 154746;  
Best Local Similarity 45.8%; Pred. No. 5.4e-05;  
Matches 259; Conservative 0; Mismatches 301; Indels 5; Gaps 1;  
QY 128 GCGCGCGCGCGCGCGCTGTGCTTCGCGGCTGCTGCGCGCTGCGCGCTGCGCGCGCA 187  
Db 129446 GCGCGCGCGCGCGCGCTGCGCGCTGCGCGCTGCGCGCTGCGCGCGCGCGCG 129505  
QY 188 GGTCTGAGCGCGAGAGGCTCAGGACCTGCGAGTCTGCGAGGAGGTCAGGAGAGCGCT 247  
Db 129506 GCGCGCGCGCGCGCGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 129565  
QY 248 CGAGCGTTCGCGCGCGCGCTCAGCGCGAGTAAACGCGCGCGCGCGCGCGCGCGCG 307  
Db 129566 GCGCGCGCGCGCGCGCGCTCAGCGCGCGCTCTTCTTCTGCGCGCGCGCGCG 129625  
QY 308 TGGCCCGTTCGCGTGGGCGACCGCGCTCCGAGTGTGCGTGTGCGAGCGCTCAGGAGCT 367  
Db 129626 GCGCGCGCTTGTGCGCGCGCGCGCTTGTGCGCTTGTGCGCGCTTGTGCGCG 129685  
QY 368 GAGCGCGAGTTCGCGTGGCGCGCGCATCCGAGAGCATGTGTGCGCGCGTACGAGGAGCGCAT 427  
Db 129686 GCGCGCGAGCGAGTTCGCGCGCGCGCGCATGTGTGCGCGAGCGAGGCGCGCGCG 129745  
QY 428 GCGCGCGTGGAGAAAGTGTGCGCGCATGCGCGCGCGCGCGCGCGCGCGCGCGCG 487  
Db 129746 TCTGGAAGAGCAGTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 129805  
QY 488 AGGAGGAGGAGGAGCGGGGCTTACTACTACCTTCGCGCGCGCGCGCGCGCGCGCG 547  
Db 129806 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 129865  
QY 548 CTACGCTGAGGTCGCGCGCGAGATGTATCAACCTGTCTGCGCGCGCGCGCGCGCG 607  
Db 129866 CTTGCTGTCGCGCGCGCGCGCGAGGT-----CCCCGCGAGCGCATCAGCACCGAGCG 129920  
QY 608 CGGGCCCAAGATCGCGCGGTGAGGCTTACGAAGGCGCGGAGTACGCCCGGGTTGCC 667  
Db 129921 GTCCGCGACGAACCGAGCTTCGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCG 129980  
QY 668 GATGATGTGCGCGCTGTGCGAGGCC 692  
Db 129981 CGGGCGCGCGCGCGCGCGCGCGCGCC 130005



```

; LOCATION: (70659)..(71906)
; OTHER INFORMATION: ORF 19; positive strandedness
; NAME/KEY: misc feature
; LOCATION: (73439)..(71964)
; OTHER INFORMATION: ORF 20; negative strandedness
; NAME/KEY: misc feature
; LOCATION: (74216)..(73563)
; OTHER INFORMATION: ORF 21; negative strandedness
; NAME/KEY: misc feature
; LOCATION: (75424)..(74213)
; OTHER INFORMATION: ORF 22; negative strandedness
; NAME/KEY: misc feature
; LOCATION: (75535)..(76464)
; OTHER INFORMATION: ORF 23; positive strandedness
; NAME/KEY: misc feature
; LOCATION: (78110)..(76449)
; OTHER INFORMATION: ORF 24; negative strandedness
; NAME/KEY: misc feature
; LOCATION: (79864)..(78107)
; OTHER INFORMATION: ORF 25; negative strandedness
; NAME/KEY: misc feature
; LOCATION: (81624)..(79861)
; OTHER INFORMATION: ORF 26; negative strandedness
; NAME/KEY: misc feature
; LOCATION: (81909)..(81682)
; OTHER INFORMATION: ORF 27; negative strandedness
; NAME/KEY: misc feature
; LOCATION: (82346)..(82062)
; OTHER INFORMATION: ORF 28; negative strandedness
; NAME/KEY: misc feature
; LOCATION: (82587)..(84446)
; OTHER INFORMATION: ORF 29; positive strandedness
; NAME/KEY: misc feature
; LOCATION: (84481)..(85548)
; OTHER INFORMATION: ORF 30; positive strandedness
; NAME/KEY: misc feature
; LOCATION: (85556)..(86845)
; OTHER INFORMATION: ORF 31; positive strandedness
; NAME/KEY: misc feature
; LOCATION: (87372)..(86803)
; OTHER INFORMATION: ORF 32; positive strandedness
; NAME/KEY: misc feature
; LOCATION: (87494)..(88420)
; OTHER INFORMATION: ORF 33; positive strandedness; N-terminus only
US-09-976-059-1

```

```

Query Match      6.6%; Score 63; DB 10; Length 88421;
Best Local Similarity 54.6%; Pred.No. 5.8e-05;
Matches 147; Conservative 0; Mismatches 120; Indels 2; Gaps 1;

QY 123 GCGCGGCGCGCGCGCGCTGTGCTTCGCGGCCCTGGTGGCGCTGCGCGTCTGCCNA 182
DB 67741 GCGGAACGAGCGCGCGCGCCCTCGCGCGCGCGCGCTGGGGCGCTGGCGACGCGCGC 67800

QY 183 GCGGAGGTGAGCGGCGAGAGGCTCAGGGACCTGCGAGTGTGCGGAGGCTCAGGAGAGC 242
DB 67801 CCGGACGACCGGTGGCGCGAGCTGCGGTGCGCGCGCGCGCGCGCGCGCGCGCGC 67860

QY 243 CCGCTCGACGCGTGGCGCGCGAGTCTTCGACCGCGCGAGCTAACCG--GCGCGCGCGCGCGCG 300
DB 67861 TGGCGCGCGCGCTGCGCAAGGTCTCTCGCCCGGTACGAGCGCGCGCGCTGCGGCTCG 67920

QY 301 GCGGGGTTGGCGCTTCGCGTGGCGCACCGGCTCGGATGCGGTCTCCAGCGAGCTCC 360
DB 67921 GCGCGCGCGCGCGGTCTGGCGCGCGCGCGCGGTGCGGTGCGCGCGCGCTCCAGCGCGC 67980

QY 361 AGGAGCTGAGCGCGAGTGCCTGCGCGC 389
DB 67981 GCTCGGGCTCGCGCGCGCGCGCGGTCCC 68009

```

RESULT 12

US-10-017-161-1483

```

; Sequence 1483, Application US/10017161
; Publication No. US2003014368A1
; GENERAL INFORMATION:
; APPLICANT: SUWA, MAKIKO
; APPLICANT: ASAI, KIYOSHI
; APPLICANT: AKIYAMA, YUTAKA
; APPLICANT: ABURATANI, HIROYUKI
; TITLE OF INVENTION: NOVEL G PROTEIN-COUPLED RECEPTORS
; FILE REFERENCE: 084335/0152
; CURRENT APPLICATION NUMBER: US/10/017,161
; CURRENT FILING DATE: 2002-12-18
; PRIOR APPLICATION NUMBER: JP 2001/246789
; PRIOR FILING DATE: 2001-06-18
; NUMBER OF SEQ ID NOS: 2430
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1483
; LENGTH: 3133
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; LOCATION: source
; FEATURE:
; LOCATION: (1)..(3133)
; NAME/KEY: CDS
; LOCATION: (201)..(2933)
; NAME/KEY: modified base
; LOCATION: (1)..(118)
; OTHER INFORMATION: a, t, c, g, unknown or other
; FEATURE:
; NAME/KEY: modified base
; LOCATION: (124)..(129)
; OTHER INFORMATION: a, t, c, g, unknown or other
; FEATURE:
; NAME/KEY: modified base
; LOCATION: (139)
; OTHER INFORMATION: a, t, c, g, unknown or other
; FEATURE:
; NAME/KEY: modified base
; LOCATION: (146)..(148)
; OTHER INFORMATION: a, t, c, g, unknown or other
; FEATURE:
; NAME/KEY: modified base
; LOCATION: (151)..(152)
; OTHER INFORMATION: a, t, c, g, unknown or other
; FEATURE:
; NAME/KEY: modified base
; LOCATION: (155)
; OTHER INFORMATION: a, t, c, g, unknown or other
; FEATURE:
; NAME/KEY: modified base
; LOCATION: (172)
; OTHER INFORMATION: a, t, c, g, unknown or other
; FEATURE:
; NAME/KEY: modified base
; LOCATION: (179)..(180)
; OTHER INFORMATION: a, t, c, g, unknown or other
; FEATURE:
; NAME/KEY: modified base
; LOCATION: (182)
; OTHER INFORMATION: a, t, c, g, unknown or other
; FEATURE:
; NAME/KEY: modified base
; LOCATION: (185)..(188)
; OTHER INFORMATION: a, t, c, g, unknown or other
; FEATURE:
; NAME/KEY: modified base
; LOCATION: (192)
; OTHER INFORMATION: a, t, c, g, unknown or other
; FEATURE:
; NAME/KEY: modified base
; LOCATION: (194)..(195)

```

```
;/ OTHER INFORMATION: a, t, c, g, unknown or other
;/ FEATURE:
;/ NAME/KEY: modified_base
;/ LOCATION: (210)
;/ OTHER INFORMATION: a, t, c, g, unknown or other
;/ FEATURE:
;/ NAME/KEY: modified_base
;/ LOCATION: (212)
;/ OTHER INFORMATION: a, t, c, g, unknown or other
;/ FEATURE:
;/ NAME/KEY: modified_base
;/ LOCATION: (214)
;/ OTHER INFORMATION: a, t, c, g, unknown or other
;/ FEATURE:
;/ NAME/KEY: modified_base
;/ LOCATION: (219)
;/ OTHER INFORMATION: a, t, c, g, unknown or other
;/ FEATURE:
;/ NAME/KEY: modified_base
;/ LOCATION: (222)..(225)
;/ OTHER INFORMATION: a, t, c, g, unknown or other
;/ FEATURE:
;/ NAME/KEY: modified_base
;/ LOCATION: (229)
;/ OTHER INFORMATION: a, t, c, g, unknown or other
;/ FEATURE:
;/ NAME/KEY: modified_base
;/ LOCATION: (231)..(232)
;/ OTHER INFORMATION: a, t, c, g, unknown or other
;/ FEATURE:
;/ NAME/KEY: modified_base
;/ LOCATION: (238)..(239)
;/ OTHER INFORMATION: a, t, c, g, unknown or other
;/ FEATURE:
;/ NAME/KEY: modified_base
;/ LOCATION: (242)
;/ OTHER INFORMATION: a, t, c, g, unknown or other
;/ FEATURE:
;/ NAME/KEY: modified_base
;/ LOCATION: (245)
;/ OTHER INFORMATION: a, t, c, g, unknown or other
;/ FEATURE:
;/ NAME/KEY: modified_base
;/ LOCATION: (248)..(260)
;/ OTHER INFORMATION: a, t, c, g, unknown or other
;/ FEATURE:
;/ NAME/KEY: modified_base
;/ LOCATION: (265)..(270)
;/ OTHER INFORMATION: a, t, c, g, unknown or other
;/ FEATURE:
;/ NAME/KEY: modified_base
;/ LOCATION: (272)..(276)
;/ OTHER INFORMATION: a, t, c, g, unknown or other
;/ FEATURE:
;/ NAME/KEY: modified_base
;/ LOCATION: (278)
;/ OTHER INFORMATION: a, t, c, g, unknown or other
;/ FEATURE:
;/ NAME/KEY: modified_base
;/ LOCATION: (282)..(283)
;/ OTHER INFORMATION: a, t, c, g, unknown or other
;/ FEATURE:
;/ NAME/KEY: modified_base
;/ LOCATION: (287)..(290)
;/ OTHER INFORMATION: a, t, c, g, unknown or other
;/ FEATURE:
;/ NAME/KEY: modified_base
;/ LOCATION: (302)..(306)
;/ OTHER INFORMATION: a, t, c, g, unknown or other
;/ FEATURE:
;/ NAME/KEY: modified_base
;/ LOCATION: (311)..(313)
;/ OTHER INFORMATION: a, t, c, g, unknown or other
```

```
;/ FEATURE:
;/ NAME/KEY: modified_base
;/ LOCATION: (323)..(334)
;/ OTHER INFORMATION: a, t, c, g, unknown or other
;/ FEATURE:
;/ NAME/KEY: modified_base
;/ LOCATION: (336)..(341)
;/ OTHER INFORMATION: a, t, c, g, unknown or other
;/ FEATURE:
;/ NAME/KEY: modified_base
;/ LOCATION: (345)..(363)
;/ OTHER INFORMATION: a, t, c, g, unknown or other
;/ FEATURE:
;/ NAME/KEY: modified_base
;/ LOCATION: (365)..(368)
;/ OTHER INFORMATION: a, t, c, g, unknown or other
;/ FEATURE:
;/ NAME/KEY: modified_base
;/ LOCATION: (370)
;/ OTHER INFORMATION: a, t, c, g, unknown or other
;/ FEATURE:
;/ NAME/KEY: modified_base
;/ LOCATION: (374)..(375)
;/ OTHER INFORMATION: a, t, c, g, unknown or other
;/ FEATURE:
;/ NAME/KEY: modified_base
;/ LOCATION: (379)..(385)
;/ OTHER INFORMATION: a, t, c, g, unknown or other
;/ FEATURE:
;/ NAME/KEY: modified_base
;/ LOCATION: (387)..(388)
;/ OTHER INFORMATION: a, t, c, g, unknown or other
;/ FEATURE:
;/ NAME/KEY: modified_base
;/ LOCATION: (390)..(393)
;/ OTHER INFORMATION: a, t, c, g, unknown or other
;/ FEATURE:
;/ NAME/KEY: modified_base
;/ LOCATION: (395)
;/ OTHER INFORMATION: a, t, c, g, unknown or other
;/ FEATURE:
;/ NAME/KEY: modified_base
;/ LOCATION: (397)
;/ OTHER INFORMATION: a, t, c, g, unknown or other
;/ FEATURE:
;/ NAME/KEY: modified_base
;/ LOCATION: (400)..(402)
;/ OTHER INFORMATION: a, t, c, g, unknown or other
;/ FEATURE:
;/ NAME/KEY: modified_base
;/ LOCATION: (409)
;/ OTHER INFORMATION: a, t, c, g, unknown or other
;/ FEATURE:
;/ NAME/KEY: modified_base
```

```
Query Match 6.4%; Score 61.2; DB 12; Length 3133;
Best Local Similarity 34.6%; Pred. No. 5.9e-05;
Matches 198; Conservative 0; Mismatches 374; Indels 0; Gaps 0;

QY 129 GCGGGCGCGCGCGCTGTGCTTCGCGGCCCTGTGTCGCGCGCGCGCGCGCGCGCGCGAG 188
DB 394 GNGGNNNGGGGGGNGGNGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 453

QY 189 GTCGAGCGCGCAGAGCTCAGGACCTGACAGTCTGTGCGAGAGGTCCAGAGAGCCCGCTC 248
DB 454 GNGGGGNGNGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 513

QY 249 GACGGGTGCCCGCAGTCTTCGACCGGCGAGCTAACCGGCGCGCGCGCGCGCGCGCTT 308
DB 514 GGGGGGNNNGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 573

QY 309 GGGCCGCTTCGGGTGGGGCACCGCGGTCCGGATGCCGTGCTGCCAGCAGCTCCAGGACGTG 368
```

[illegible]

RESULT 13  
US-10-116-275-197  
; Sequence 197, Application US/10116275  
; Publication No. US20030211476A1  
; GENERAL INFORMATION:  
; APPLICANT: Elan Pharmaceutical Technology  
; APPLICANT: O'Mahony, Daniel J.  
; APPLICANT: Brayden, David  
; APPLICANT: Byrne, Daragh  
; APPLICANT: Lambkin, Imelda  
; APPLICANT: Higgins, Lisa  
; TITLE OF INVENTION: Genetic Analysis of Peyer's Patches and M Cells and Methods and  
; FILE REFERENCE: E1067/20087  
; CURRENT APPLICATION NUMBER: US/10/116,275  
; CURRENT FILING DATE: 2002-10-04  
; NUMBER OF SEQ ID NOS: 349  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 197  
; LENGTH: 3824  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-10-116-275-197

Query Match	6.3%;	Score 59.4;	DB 12;	Length 3824;
Best Local Similarity	48.8%;	Pred. No. 0.00017;		
Matches 189;	Conservative 0;	Mismatches 196;	Indels 2;	Gaps 1;
Qy	145	TGTGCTTCGGGGCCCTGTGSCCTGTGCGCCCTCTCCAAAGGCGAGGTCCGAGCGGCAGAGGC	204	
Db	593	TCTCTTCGCGCTCGCTTCGCGCTCATGGCCGCGCGGGCGCGGGCGCGCGCGCGCGCGCG	652	
Qy	205	TCAGGGACCTGCAGTGTGGCAGGAGGTCCAGGAGAGCCCTCTCGACGCGTGTCCGCCAGG	264	
Db	653	CGCGCGCGCCACGACGGCCCGGGGGCGGTGTGCGGCCCGCGCGCGCGCGCGCGCG	712	
Qy	265	TCCTCGACCGGCAGCTAACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG	324	
Db	713	CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG	772	
Qy	325	GCACCGGGCTCCGGATGCGGTGTGTCAGACAGCTCCAGGACGTTCAGCGCGCGAGTGCCT	384	
Db	773	CGCGCGGGCTCCTGGCGGGCTTCGCGGACCCCTCACCCGCGATATGCACAGCTGGGCGCAC	832	
Qy	385	CGCGCGGCCATCCGGAGCATGCTCAGGGGCTTACGAGGAGGCCATCCCGCGCTGGAGAAAG	444	

Db	833	TGTGCGACCCGGCGCGCGCGCCCATGAACATGCTCCGTCGGCGCTGCCGC	--ACCCCGG	890
Qy	445	GCTGCTGGCCATGCGGGCGCGCAGCAGCAGCCGCGCGCGAGGAGGAGGACACAGG	504	
Db	891	GCTGCTGGCGCGCGCGCGCGCACACACGCGCGGCGCAGCGCGCGCGGCGT	CGGCCGG	950
Qy	505	GGGGCTACTACTACCCCTCTGCAGCGCGC	531	
Db	951	GCAGTGGCAGCGGCATCGCGCGCGC	977	

RESULT 14

US-09-864-864-311/c

; Sequence 311, Application US/09864864

; Patent NO. US20020102679A1

; GENERAL INFORMATION:

; APPLICANT: Xu, Jiangchun

; APPLICANT: Mitcham, Jennifer L.

; APPLICANT: Harlocker, Susan L.

; APPLICANT: Dillon, Davin C.

; APPLICANT: Secrist, Heather

; APPLICANT: Lodes, Michael J.

; APPLICANT: Algate, Paul A.

; APPLICANT: Fling, Steve P.

; APPLICANT: Mannion, Jane

; APPLICANT: Benson, Darin R.

; APPLICANT: Carter, Darrick

; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY

; TITLE OF INVENTION: AND DIAGNOSIS OF OVARIAN CANCER

; FILE REFERENCE: 210121.523

; CURRENT APPLICATION NUMBER: US/09/864, 864

; CURRENT FILING DATE: 2001-05-23

; NUMBER OF SEQ ID NOS: 341

; SOFTWARE: Corixa Invention Disclosure Database

; SEQ ID NO 311

; LENGTH: 3035

; TYPE: DNA

; ORGANISM: Homo sapiens

US-09-864-864-311

[illegible]

RESULT 15  
US-09-814-353-19995/c  
; Sequence 19995, Application US/09814353  
; Publication No. US20030165631A1  
; GENERAL INFORMATION:  
; APPLICANT: Lee, John  
; APPLICANT: Thompson, Pamela  
; APPLICANT: Lillie, James

; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR  
; TITLE OF INVENTION: IDENTIFICATION, ASSESSMENT, PREVENTION, AND  
; TITLE OF INVENTION: THERAPY OF OVARIAN CANCER

; FILE REFERENCE: MRI-006B  
; CURRENT APPLICATION NUMBER: US/09/814,353  
; CURRENT FILING DATE: 2001-03-21  
; PRIOR APPLICATION NUMBER: US 60/191,031  
; PRIOR FILING DATE: 2000-03-21  
; PRIOR APPLICATION NUMBER: US 60/207,124  
; PRIOR FILING DATE: 2000-05-25  
; PRIOR APPLICATION NUMBER: US 60/211,940  
; PRIOR FILING DATE: 2000-06-15  
; PRIOR APPLICATION NUMBER: US 60/216,820  
; PRIOR FILING DATE: 2000-07-07  
; PRIOR APPLICATION NUMBER: US 60/220,661  
; PRIOR FILING DATE: 2000-07-25  
; PRIOR APPLICATION NUMBER: US 60/257,672  
; PRIOR FILING DATE: 2000-12-21  
; NUMBER OF SEQ ID NOS: 22037  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 19995  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-814-353-19995

Query Match 6.2%; Score 59; DB 12; Length 3393;  
Best Local Similarity 52.2%; Pred. No. 0.00021;  
Matches 131; Conservative 0; Mismatches 120; Indels 0; Gaps 0;  
Qy 290 CGCGCGCGCGCGCGGCGTTGCGCGTTCGCGTGGGCGACCGGGCTCCGGATGCGGTGCTG 349  
Db 389 CGCGCGCGCGCGGAGCCCTCCCGCGCGCGGCTGCCAGAGCGCGCGAGCGCGCTGCTG 330  
Qy 350 CCAGCAGCTCCAGGACGTGAGCGCGAGTCCGCTGCGCGCCATCCGGAGCATGGTTCAG 409  
Db 329 CCGAGGCTGCGCGCGGAGCGCGCCATTACACGCGCGCGGGGGCGCCAGGGGGGAG 270  
Qy 410 GGGCTAGGAGGAGCCATCGCGCGTGGAGAAAGGTGGTGGCCATGGGGGGCGCAGCA 469  
Db 269 GAGGGGCGCGGGCGCGCGCGCGCGCGCAGCGCGCTCGCGCCCAACTGCTGGGCGCGT 210  
Qy 470 GCAGCGCGCGCGCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 529  
Db 209 GCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 150  
Qy 530 GCCAGGAGAGG 540  
Db 149 CGCGCGGACG 139

Search completed: November 29, 2003, 17:04:01  
Job time : 371 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 29, 2003, 15:53:04 ; Search time 58 Seconds  
(without alignments)  
563.753 Million cell updates/sec

Title: US-10-053-410-4

Perfect score: 1148

Sequence: 1 MAKIAAAAAALCFALVAV.....MMCRLEPQECIFSGGDY 206

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A Geneseq\_19Jun03.\*

```
1: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:*
2: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:*
3: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:*
4: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:*
5: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:*
6: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:*
7: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:*
8: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:*
9: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:*
10: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:*
11: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:*
12: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:*
13: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:*
14: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:*
15: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:*
16: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:*
17: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:*
18: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:*
19: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:*
20: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:*
21: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:*
22: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:*
23: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:*
24: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2003.DAT:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	156	13.6	434	23	AAW50386
2	130.5	11.4	157	19	AAW53261
3	117	10.2	362	22	AB362882
4	114.5	10.0	143	13	AA328122
5	110	9.6	385	22	AB371788
6	110	9.6	797	22	AB364072
7	109.5	9.5	426	14	AA331036
8	109	9.5	166	24	ABJ26643
9	106.5	9.3	174	22	AAV72901

10	106.5	9.3	1145	22	AAE09770	Drosophila melanog
11	105	9.1	694	23	AAU74823	Human REPR 6 prot
12	105	9.1	694	24	AAE34057	F2D 8 protein. Un
13	105	9.1	694	24	ABU55903	Human protein Friz
14	103.5	9.0	348	22	AB559015	Drosophila melanog
15	103	9.0	202	22	AB62974	Drosophila melanog
16	103	9.0	655	23	AB57126	Mouse ischaemic co
17	101	8.8	217	20	AAW97318	A HOQ059 polypept
18	101	8.8	217	20	AAW97318	Human polypeptide
19	101	8.8	228	22	AAW41673	Human polypeptide
20	101	8.8	515	21	AB335806	Protein involved i
21	101	8.8	894	22	AB335148	Human NFAR-2 SEQ I
22	100.5	8.8	360	23	ABP63025	Human polypeptide
23	100.5	8.8	360	24	ABP97405	Human embryonic st
24	100	8.7	656	14	AA44555	Human EMS protein
25	100	8.7	656	19	AAW33813	Human EMS protein
26	100	8.7	665	22	ABG06460	Novel human diagno
27	100	8.7	926	22	AB65135	Drosophila melanog
28	99.5	8.7	134	22	AAO05578	Human polypeptide
29	98.5	8.6	124	15	AA58636	Amylase inhibitor
30	98.5	8.6	124	18	AAW29525	Wheat amylase inhi
31	98.5	8.6	124	18	AAW24493	Novel amylase inhi
32	98	8.5	169	21	AB28185	Human heterogeneou
33	97.5	8.5	302	23	ABP63092	FLO11 gene expres
34	97.5	8.5	336	22	ABG12614	Novel human diagno
35	97.5	8.5	474	21	AAW27305	Arabidopsis thalia
36	97.5	8.5	477	21	AAW27304	Arabidopsis thalia
37	97.5	8.5	533	21	AAW27303	Arabidopsis thalia
38	97.5	8.5	671	22	ABG20586	Novel human diagno
39	97	8.4	144	21	AAW84357	Amino acid sequenc
40	97	8.4	148	15	AAW47481	Polypeptide encode
41	97	8.4	297	24	ABR41261	Human DITP4 intrac
42	96.5	8.4	207	22	AB82383	Peanut allergen Ar
43	96	8.4	140	10	AAW91891	Part of the sequen
44	96	8.4	173	21	AAW84354	Amino acid sequenc
45	96	8.4	454	22	AB61917	Drosophila melanog

#### ALIGNMENTS

#### RESULT 1

AAW50386  
ID AAW50386 standard; Protein; 434 AA.

AC AAW50386;

DT 18-FEB-2002 (first entry)

DE Wheat glutenin variant 1Ax2asteriskB.

DE Wheat; glutenin; storage protein; variant; transgenic plant;  
gluten.

OS Triticum aestivum.

PN WO200179477-A2.

PD 25-OCT-2001.

PF 13-APR-2001; 2001WO-HU00045.

PR 14-APR-2000; 2000HU-0001563.

PA (MAGY ) MTA MEZOGAZDASAGI KI.

PI Juhasz A., Tamas L., Bedo Z., Vida G., Karsai I., Lang L;

PI Tamasne Nyitrai E;

XX WPI; 2002-017612/02.

DR N-PSDB; AA170852.

PT A new nucleic acid comprising a variant allele of the 1Ax2asteriskHMW

PT Glutenin subunit gene of wheat has at least one extra cysteine codon  
PT and is useful for developing new wheat varieties with enhanced quality  
PT gluten

XX Disclosure; Fig 6; 28pp; English.

XX The present sequence is that of the newly identified allele,  
CC designated 1A2asteriskB, of 1A2asterisk HMW glutenin, found in  
CC certain lines of wheat variety Bankuti 1201. The variant gene  
CC (see AA170852) differs from the 1A2asterisk gene by a single point  
CC mutation of C to G. This results in a change from Ser to Cys at  
CC amino acid position 394 of the encoded protein. Because of this,  
CC the newly identified allele provides enhanced possibilities for the  
CC formation of disulfide bridges compared with the known allele,  
CC which could at least partly explain the good technological  
CC properties of the variety. The newly identified gene can be used  
CC to develop transgenic wheat varieties with better quality gluten.

XX Sequence 434 AA;

Query Match 13.6%; Score 156; DB 23; Length 434;  
Best Local Similarity 24.8%; Pred. No. 2.5e-06;  
Matches 62; Conservative 17; Mismatches 43; Indels 128; Gaps 11;

QY 11 ALCFAALVAVACQGEVERQRLDLOCWQVQESPLDACRQLVDRQLTGCGGGGVGPPR 70  
DB 9 AAVVVALVALTAAGEASGO---LQCELEQSHSLKACRQVVD----- 48  
QY 71 WGTGLMRCCQQLQDVSRRCRAIRSMVRGYEE--AMPPLKGMWP----- 115  
DB 49 -----QQLRDVSPQCQVGGGFVARQYEQVVVPPKGSFPGTTPPQQLQOSI 98  
QY 116 -WG-----RQQP-----PPQG-----GGQGG 133  
DB 99 LWGIPALLRRYLSVTSPQVYVPGQASSRPGQEQEYLLTSPQSGQWQPGQSG 158  
QY 134 Y-----YFPCS-----RPGEGYGGG-----GQRQ 154  
DB 159 YVTPSPQSGQKQPGYVTPSPWQPEQLQPTQQRQPGQQLRQGGQSGQGP 218  
QY 155 MYPPC--RPG 162  
DB 219 YVTPSSQFG 228

RESULT 2

AAW53261  
ID AAW53261 standard; Protein; 157 AA.

XX AAW53261;

XX 20-JUL-1998 (first entry)

XX Amino acid sequence of Synthetic HMW-glutenin protein.

XX HMW glutenin subunit repeat; wheat; synthetic; repetitive domain;  
KW viscoelastic property; wheat flour; wheat dough; bread; noodle.

XX Synthetic.

XX Triticum sp.

XX WO9808607-A1.

XX 05-MAR-1998.

XX 29-AUG-1997; 97WO-US15398.

XX 30-AUG-1996; 96US-0706391.

XX (USDA ) US SEC OF AGRIC.

XX Anderson OD;

DR WPI; 1998-179218/16.

XX A process for altering the viscoelasticity of a dough - comprises  
PT genetic engineering of a high molecular weight glutenin subunit  
PT containing a non-natural repetitive domain

XX Example 2; Page 18; 28pp; English.

XX This amino acid sequence is of the synthetic HMW-glutenin protein  
CC which is introduced into the DNA of the cereal to form a non-natural  
CC repetitive domain which changes the viscoelastic property of a dough.  
CC The domain comprises HMW glutenin subunit repeats which are synthetic  
CC in sequence or non-natural in number. The dough is useful for making  
CC a wheat flour, wheat dough, bread or noodle.

XX Sequence 157 AA;

Query Match 11.4%; Score 130.5; DB 19; Length 157;  
Best Local Similarity 27.2%; Pred. No. 0.0002;  
Matches 55; Conservative 21; Mismatches 53; Indels 73; Gaps 11;

QY 24 QGEVERQRLDLOC---WQVQESPLDACRQLVDRQLTGCGGGGVGPPRMTGLMRCC 80  
DB 2 EGEASEQ---LQCELEQSHSLKACRQVVD----- 31  
QY 81 QQLQDVSRRCRAIRSMVRGYEE--AMPPLKGMWPGRQRPQPPQ-----GGGGGG 132  
DB 32 QQLRDVSPQCQVGGGFVARQYEQVVVPPKGSFPGTTPPQQLQOSI 89  
QY 133 GYYP---CSRPGEGYGGGQGRQWYP---PCRPGTT---GGPRIGRVLTKAREVA 182  
DB 90 KRYFSVTCQ-----QVSYVPGQASFORSSSYHVSVEHQAAASLKAKAQAQLA 138  
QY 183 AGLPMMCRILSEFQECISFSGD 204  
DB 139 AQLPAMCRL-----EGGD 151

RESULT 3

ABB62882

ID ABB62882 standard; Protein; 362 AA.

XX ABB62882;

XX 26-MAR-2002 (first entry)

XX Drosophila melanogaster polypeptide SEQ ID NO 15438.

XX Drosophila; developmental biology; cell signalling; insecticide;  
KW pharmaceutical.

XX Drosophila melanogaster.

XX WO200171042-A2.

XX 27-SEP-2001.

XX 23-MAR-2001; 2001WO-US09231.

XX 23-MAR-2000; 2000US-191637P.

XX 11-JUL-2000; 2000US-0614150.

XX (PEKE ) PE CORP NY.

XX Venter JC, Adams M, Li PWD, Myers EW;

XX WPI; 2001-656860/75.

XX N-PSDB; ABL06985.

XX New isolated nucleic acid detection reagent for detecting 1000 or more  
PT genes from Drosophila and for elucidating cell signalling and cell-cell  
PT interactions -



Query Match 9.6%; Score 110; DB 22; Length 385;  
Best Local Similarity 29.7%; Pred. No. 0.046;  
Matches 41; Conservative 7; Mismatches 50; Indels 40; Gaps 5;

QY 35 LQWQVQESPLDA-----CRQLDRQLTGGGGGGVGPFRWGTGLMRCCQLQDVSRREC 90  
DB 174 LQTHSIKNTLDVKAIKQDMDRQ---GGGGGGGPRAGGRGQ-----216

QY 91 RCAAIRSMVRGYEAMPPEKGNWPHGRQOQPPPGGGGGGGVYPCSRPGEYGYGOG 150  
DB 217 ---GDRGQGGG-----GWGGQNRQNGGNGGAGGGGGFGNSGGNFGGGQGGSG 263

QY 151 GQRQYPPCRPTGTGGP 168  
DB 264 GWNQ-----GGSGGP 275

RESULT 6  
ABB64072  
ID ABB64072 standard; Protein; 797 AA.  
XX  
AC ABB64072;  
XX  
DT 26-MAR-2002 (first entry)  
XX  
DE Drosophila melanogaster polypeptide SEQ ID NO 19008.  
XX  
KW Drosophila; developmental biology; cell signalling; insecticide;  
KW pharmaceutical.  
XX  
OS Drosophila melanogaster.  
XX  
PN WC200171042-A2.  
XX  
PD 27-SEP-2001.  
XX  
PF 23-MAR-2001; 2001WO-US09231.  
XX  
PR 23-MAR-2000; 2000US-191637P.  
PR 11-JUL-2000; 2000US-0614150.  
XX  
PA (PEKE ) PE CORP NY.  
XX  
PI Venter JC, Adams M, Li PWD, Myers EW;  
XX  
DR WPI; 2001-656860/75.  
DR N-PSDB; ABL08175.  
XX  
XX New isolated nucleic acid detection reagent for detecting 1000 or more  
genes from Drosophila and for elucidating cell signalling and cell-cell  
interactions -  
XX  
PS Disclosure; SEQ ID NO 19008; 21pp + Sequence Listing; English.  
XX  
CC The invention relates to an isolated nucleic acid detection reagent  
capable of detecting 1000 or more genes from Drosophila. The invention is  
useful in developmental biology and in elucidating cell signalling and  
cell-cell interactions in higher eukaryotes for the development of  
insecticides, therapeutics and pharmaceutical drugs. The invention  
discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA  
sequences (ABL01840-ABL16175) and the encoded proteins  
(ABBS7737-ABBS7202).  
CC  
CC The sequence data for this patent did not form part of the printed  
specification, but was obtained in electronic format directly from WIPO  
at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
XX  
SQ Sequence 797 AA;

Query Match 9.6%; Score 110; DB 22; Length 797;  
Best Local Similarity 30.1%; Pred. No. 0.1;  
Matches 41; Conservative 14; Mismatches 57; Indels 24; Gaps 7;

QY 38 WQEVQES-----PLDACRQVL-DRQLTGGGGGGV-GPFRWGTGLMRCCQLQDVSRREC 90

DB 658 WAERYRSVGKIEAEIAENTLANKQNGSGGSTPNPQGGSGQ-----QQPNAAAAA 712

QY 91 RCAAIRSMVRGYEAMPPEKGNWPH-----GROQPPPGGGGGGGG---YYPCSR 140

DB 713 AAAAAAGGAGGQSWTPTQYAYQYQYAAAAAGGQPGAPQPGGGGGGPPGNY 772

QY 141 PGEYGV--YGQGGGQ 154

DB 773 PGAGYGGYPAFGQQQ 788

RESULT 7  
AAR31036  
ID AAR31036 standard; Protein; 426 AA.  
XX  
AC AAR31036;  
XX  
DT 25-MAR-2003 (updated)  
DT 26-MAY-1993 (first entry)  
XX  
DE C. acidovorans stereospecific hydrolase.  
XX  
KW S-(+)-2,2-dimethylcyclopropane carboxamide; 2,2-DMCPCA; prodn.;  
cilastatin; antibiotic therapy; penem; carbapenem.  
XX  
OS Comamonas acidovorans A:18.  
XX  
FH Key Location/Qualifiers  
FT Peptide 1..7  
FT Peptide /note= "N-terminal peptide used to obtain AAR35116"  
FT Peptide 13..18  
FT Peptide /note= "peptide used to obtain AAR35117"

XX  
PN EP524604-A2.  
XX  
PD 27-JAN-1993.  
XX  
PF 21-JUL-1992; 92EP-0112446.  
XX  
PR 26-JUL-1991; 91CH-0002247.  
XX  
PA (LONZ ) LONZA AG.  
XX  
PI Zimmermann T, Robins K, Birch OM, Boehlen E;  
XX  
DR WPI; 1993-028686/04.  
DR N-PSDB; AAR38949.  
XX  
PT Pure S-(+)-2,2-dimethylcyclopropane carboxamide prodn. by  
microorganisms (S) - with a stereospecific hydrolase isolated from  
PT microorganism, useful for inhibition of renal dehydropeptidase  
via cilastatin  
XX  
PS Disclosure; Fig 3; 19pp; German.  
XX  
CC The sequence is that of Comamonas acidovorans A:18 stereospecific  
hydrolase which can be used in the prodn. of optically pure S-(+)-2,2-  
dimethylcyclopropane carboxamide (2,2-DMCPCA), the starting  
material for the prodn. of cilastatin. Cilastatin inhibits e.g. renal  
dehydropeptidase (RDP) and in therapy it is administered with the  
CC antibiotic penem or carbapenem to prevent their inactivation by RDP.  
CC (Updated on 25-MAR-2003 to correct PN field.)  
CC (Updated on 28-MAR-2003 to correct PI field.)  
XX  
SQ Sequence 426 AA;

Query Match 9.5%; Score 109.5; DB 14; Length 426;  
Best Local Similarity 22.8%; Pred. No. 0.057;  
Matches 64; Conservative 16; Mismatches 70; Indels 131; Gaps 12;

QY 24 QGEVERQRRLDLCWQEVQESPLD--ACRQLDRQLTGGGGGGV-----66

```
Db 123 QAAADRRRLRRPPSRDHCPRQPLERFAMARGLVORLGRHGCGAVLRAGHGRGLHPLS 182
QY 67 -----GPFRTGTGLRNRCCQLODVSRCCRCAAIRSM 98
Db 183 IGRQRHHGAQHLGQGEPPRLRTGRVPGHR -----PDGAQCCRC---RSH 226
QY 99 VRGYEAMP-----PLEK-----GWMP---WGRQQQPPPGGGGGQGG 133
Db 227 ARGHRRGGGAGPYGQPVORARLSCHDDARLLRPAPGHPAMGTGRRCCLPFGGGAGPGG 286
QY 134 YYPCSRPBGEGYGGGQGRQMP--PCRPTTGGG----- 167
Db 287 -----GAAPGQGRAGGPLSRCHFG--GGGLAGAVRGDRRGARRHVPCTARGL 332
QY 168 -PRIGRVLRTKAREYAAGLPMWCR-----LSEPOECSIFSGG 203
Db 333 WPARVRDRPGAGPVHRHLPAAAAAPRGLHGPACTLRAGG 373

RESULT 8
ABJ26643
ID ABJ26643 standard; Protein; 166 AA.
XX
AC ABJ26643;
XX
XX 17-APR-2003 (first entry)
XX
DE Rice seed allergen protein - SEQ ID No 48.
XX
KW Rice; Chrohn's disease antibody-binding peptide; Crohn's disease;
KW human vacuolar H+ transport ATPase; rice allergen;
KW human kruppel-like zinc finger protein 300.
XX
OS Oryza sativa.
XX
XX WO200288175-A1.
XX
XX 07-NOV-2002.
XX
XX 24-APR-2002; 2002WO-JP04061.
XX
XX 24-APR-2001; 2001JP-0126121.
XX
XX 25-FEB-2002; 2002JP-0047384.
XX
XX (SAKA ) OTSUKA PHARM CO LTD.
XX
XX Saito H, Katsuragi K, Tachikawa T, Tanaka M, Ogino K, Taki T;
XX
XX WPI; 2003-140169/13.
XX
XX Crohn's disease antibody-binding peptide applicable in examination
XX reagents for use in diagnosis of Crohn's disease
XX
XX Claim 8; Page 111; 120pp; Japanese.
XX
XX The invention comprises Chrohn's disease antibody-binding peptides. The
XX peptides of the invention are useful for examining and diagnosing Crohn's
XX disease by recognising a human vacuolar H+ transport ATPase, or human
XX kruppel-like zinc finger protein 300, or rice allergen in a biological
XX sample from a subject. The present amino acid sequence represents a rice
XX polypeptide used in the invention.
XX
XX Sequence 166 AA;
XX
XX Query Match 9.5%; Score 109; DB 24; Length 166;
XX Best Local Similarity 26.0%; Pred. No. 0.022;
XX Matches 39; Conservative 11; Mismatches 44; Indels 56; Gaps 6;
XX
XX 46 LDACROVLDRQLTGGGGGGVGRFRTGLRMWCCQLODVSR--CRCAAIRSMVRGYEB 104
Db 52 LPRCAVVRKRCQCGHGAAGPAGVDEQ-----LRQDCRQLAAVDSDSWCRCSALNHV----- 102
QY 105 AMPLEKGNWPMCRQPPPGGGGGGGGGYFCSPRPGEGYGGGQGRQMYPPCRPGTT 164
```

```
Db 103 -----GGIYRELGATDVGH-----PMAEVFGCRGD- 129
QY 165 GGGPRIGRVLRTKAREYAAGLPMWCRSELSEP 194
Db 130 -----LERA---AASLPAFNCVDIP 146

RESULT 9
AAV72901
ID AAV72901 standard; Protein; 174 AA.
XX
AC AAV72901;
XX
XX 31-MAY-2001 (first entry)
XX
DE Flax 2S storage protein.
XX
XX Flax; seed-specific promoter; storage protein; seed oil; vaccine;
KW protein expression; anticoagulant; cytokine; growth factor; pectinase;
KW interleukin; alpha-1-antitrypsin; anti-obesity protein; haemoglobin;
KW serum albumin; insulin; lactoferrin; myoglobin; pulmonary surfactant;
KW alpha amylase.
XX
XX Linum usitatissimum.
XX
XX Key Location/Qualifiers
XX Region 126..136
XX /note= "Glutamine-rich region"
XX
XX WO200116340-A1.
XX
XX 08-MAR-2001.
XX
XX 25-AUG-2000; 2000WO-CA00988.
XX
XX 27-AUG-1999; 99US-0151044.
XX
XX 27-OCT-1999; 99US-0161722.
XX
XX 30-MAY-2000; 2000CA-2310304.
XX
XX (SEMB-) SEMBIOSYS GENETICS INC.
XX (CSIR ) COMMONWEALTH SCI & IND RES ORG.
XX
XX Chaudhary S, Van Rooijen G, Moloney MM, Singh S;
XX
XX WPI; 2001-226693/23.
XX
XX N-PSDB; AAD03029.
XX
XX Expressing non-native genes in flax seeds and seeds of other plant
XX species for altering the seed oil and protein composition in the seeds,
XX comprises using seed-specific promoters obtained from flax
XX
XX Claim 14; Fig 3; 68pp; English.
XX
XX The present sequence is a flax 2S storage protein.
XX
XX The present invention relates to a method for expressing non-native genes
XX in flax seeds. The method comprises introducing a chimeric nucleic acid
XX construct containing a seed-specific promoter obtained from flax and a
XX nucleic acid which is non-native to the promoter, into a flax plant cell
XX and growing the plant cell into a mature flax plant capable of setting
XX seed, where the nucleic acid is expressed in the seed under the control
XX of the promoter. The seed-specific promoters obtained from flax are
XX useful for modifying the protein, oil or polysaccharide composition of
XX the flax seeds and seeds of other plant species. The promoters facilitate
XX expression of proteins, including sulphur-rich protein that are found in
XX lupine or Brazil nuts in a seed deficient in sulphurous amino acids,
XX peptides having pharmaceutical value such as anticoagulants, antibodies,
XX vaccines, cytokines, growth factors, interleukins, mammalian proteins,
XX including alpha-1-antitrypsin, anti-obesity proteins, haemoglobin, blood
XX proteins, human serum albumin, insulin, lactoferrin, myoglobin, pulmonary
XX surfactants and proteins of industrial value such as alpha-amylase,
XX arabinase, amyloglucosidase, catalase, cellobiohydrolase, pectinases,
XX phytase, papain and xylanase.
```



PI Griffin JA, Kallick DA, Tribouley CM, Yue H, Nguyen DB, Tang YT;  
PI Lal P, Policky JL, Azimzai Y, Lu DAM, Graul R, Yao MG, Burford N;  
PI Hafalia AJA, Baughn MR, Bandman O, Patterson C, Yang J, Xu Y;  
PI Gandhi AR, Warren BA, Ding L, Sanjanwala MS, Duggan BM, Lu Y;  
XX WPI; 2002-090432/12.  
DR N-PSDB; ABK15174.  
XX  
XX Twelve human receptors (referred to as REPTR-1 to REPTR-12), useful in  
PT the diagnosis, treatment and prevention of gastrointestinal (e.g.  
PT gastritis), autoimmune/inflammatory (e.g. osteoarthritis) and cell  
PT proliferative (e.g. cancer) disorders -  
XX  
PS Claim 50; Page 119-121; 157pp; English.  
XX  
XX This invention relates to twelve human receptors cDNA sequences  
CC referred to as REPTR-1 to REPTR-12), and the proteins encoded thereby.  
CC The proteins of the invention may have antiinflammatory, cytostatic,  
CC immunosuppressive, antiviral, anti-HIV, antiarthritic, muscular active  
CC general, anticonvulsant, nootropic, neuroprotective, anti-allergic  
CC activities. The sequences of the invention may be used to produce REPTR  
CC agonists or antagonists, and the protein sequences may be used to raise  
CC anti-REPTR antibodies. These molecules and the REPTR polynucleotides and  
CC polypeptides of the invention are useful in the diagnosis, treatment and  
CC prevention of gastrointestinal (e.g. gastritis, oesophageal carcinoma,  
CC Crohn's disease, irritable bowel syndrome, ulcerative colitis),  
CC endocrine (e.g. hypothyroidism, Kallman's disease), autoimmune/  
CC inflammatory (e.g. acquired immune deficiency syndrome (AIDS),  
CC rheumatoid arthritis, allergies, osteoarthritis, diabetes mellitus,  
CC multiple sclerosis, systemic lupus erythematosus), cell proliferative  
CC (e.g. cancer), developmental (e.g. Duchenne and Becker muscular  
CC dystrophy), neurological (e.g. epilepsy, Alzheimer's disease,  
CC Huntington's disease) and reproductive (e.g. infertility, endometriosis)  
CC disorders. Numerous other examples of each disorder are given in the  
CC specification. The present sequence represents the human REPTR6 protein  
XX sequence of the invention.  
XX  
SQ Sequence 694 AA;

Query Match 9.1%; Score 105; DB 23; Length 694;  
Best Local Similarity 23.7%; Pred. No. 0.26;  
Matches 51; Conservative 22; Mismatches 72; Indels 70; Gaps 10;  
QY 5 AAAAAALCFALVAVAVCOG-----EVER-QRLRDLCQWQEV 41  
DB 25 AAAAAKELACQEIITVPLCKGIGYNTYMPNQFNHDTQDEAGLEVHQFWPLVEIQSPDL 84  
QY 42 Q-----ESPLDACRQVLDRLQTLGGGGGGVGP-----FRWGTGLMRCC 80  
DB 85 KFFLCSMYTPICLEDYKPLPPCRSVCE-----AKAGCAPLMROYGFANPD--RMRC 136  
QY 81 QQLQDVSRERC-----AAIRSMVRGYEAMPLEKGMWPGWGRQQQPPPGGGGGGGY 135  
DB 137 RLPEQGNPDTLCDYNRDITLTTAAPSPPRLPPPPPG-----EQPPSGSGHGRP--- 185  
QY 136 YPCSR-PGEGYGYGGGQGMYPCCPGTGGGPR 169  
DB 186 -FGARPPHRRGGGGGGGDAAPPARGGGGGGKAR 219

RESULT 12  
ID AAE34057  
XX AAE34057 standard; Protein; 694 AA.  
AC AAE34057;  
XX  
DT 02-MAY-2003 (first entry)  
DE FZD 8 protein.  
XX  
XX Drug screening; toxicology assay; signalling pathway; FZD 8.  
XX Unidentified.

XX Key Location/Qualifiers  
FH Misc-difference 434  
FT /note= "Encoded by ATC"  
XX  
PN WO2002090992-A2.  
XX  
PD 14-NOV-2002.  
XX  
XX 29-APR-2002; 2002WO-CB01946.  
XX  
XX 04-MAY-2001; 2001GB-0011004.  
XX  
PA (AXOR-) AXORDIA LTD.  
XX  
PI Andrews P, Draper J, Walsh J;  
XX  
DR WPI; 2003-120579/11.  
DR N-PSDB; AAD52555.  
XX  
XX Identifying biologically active agents comprises cloning transfected  
PT cells into a cell array, exposing the array to an agent to be tested,  
PT and detecting signals generated by a reporter molecule as a result of  
PT exposure to the agent -  
XX  
PS Claim 16; Fig 66; 90pp; English.  
XX  
XX The present invention relates to a novel screening method which enables  
CC the identification of biologically active agents which mediate their  
CC effect through the activation of genes. The method involves providing a  
CC population of cells stably transfected with a nucleic acid encoding a  
CC reporter molecule, cloning the transfected cells into a cell array,  
CC exposing the array to at least one agent to be tested and detecting a  
CC signal generated by the reporter molecule as a result of exposure to  
CC the agent. The method is useful in identifying biologically active agents  
CC and the genes through which the agents act, in screening potential drugs  
CC for their ability to activate certain drug targets in a high-throughput  
CC assay, in identifying relationships between signalling pathways and  
CC specific signals that could be useful in eventually directing the  
CC differentiation of embryonic stem cells and in toxicology assays by  
CC testing for unwanted activation or inhibition of specific signalling  
CC pathways. The present sequence is FZD 8 protein used to illustrate the  
CC method of the invention.  
XX  
SQ Sequence 694 AA;  
Query Match 9.1%; Score 105; DB 24; Length 694;  
Best Local Similarity 23.7%; Pred. No. 0.26;  
Matches 51; Conservative 22; Mismatches 72; Indels 70; Gaps 10;  
QY 5 AAAAAALCFALVAVAVCOG-----EVER-QRLRDLCQWQEV 41  
DB 25 AAAAAKELACQEIITVPLCKGIGYNTYMPNQFNHDTQDEAGLEVHQFWPLVEIQSPDL 84  
QY 42 Q-----ESPLDACRQVLDRLQTLGGGGGGVGP-----FRWGTGLMRCC 80  
DB 85 KFFLCSMYTPICLEDYKPLPPCRSVCE-----AKAGCAPLMROYGFANPD--RMRC 136  
QY 81 QQLQDVSRERC-----AAIRSMVRGYEAMPLEKGMWPGWGRQQQPPPGGGGGGGY 135  
DB 137 RLPEQGNPDTLCDYNRDITLTTAAPSPPRLPPPPPG-----EQPPSGSGHGRP--- 185  
QY 136 YPCSR-PGEGYGYGGGQGMYPCCPGTGGGPR 169  
DB 186 -FGARPPHRRGGGGGGGDAAPPARGGGGGGKAR 219

RESULT 13  
ID ABUS5903  
XX ABUS5903 standard; Protein; 694 AA.  
AC ABUS5903;  
XX





CC The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 348 AA;

Query Match 9.0%; Score 103.5; DB 22; Length 348;  
Best Local Similarity 25.4%; Pred. No. 0.17;  
Matches 36; Conservative 11; Mismatches 38; Indels 57; Gaps 4;

QY 48 ACRQVLDRLTGGGGGGVGPFRWGTLRMRCQQQLQDVSRRCRCAAIRSMVRGYEAMP 107  
Db 13 ACNATFLSLGGGGGGGGGSK-----TTYNVIAT 43

QY 108 PLEKGMWPMGRQQPPPGGGGGGGGGYPCSRPG-----EGYGYGGGGGQRMYPPCR 160  
Db 44 PSSGGGGGG-----GGGGGGGGHGYSAQGGGGGGHGYAQGHGYGHG----- 85

QY 161 PGTGGGPRIGRVLTKAREYA 182  
Db 86 ---HGGSPQIIKVLQEGGGYS 104

RESULT 15  
ABB62974  
ID ABB62974 standard; Protein; 202 AA.  
XX  
AC ABB62974;  
XX  
DT 26-MAR-2002 (first entry)  
XX  
DE Drosophila melanogaster polypeptide SEQ ID NO 15714.  
XX  
KW Drosophila; developmental biology; cell signalling; insecticide;  
KW pharmaceutical.  
XX  
OS Drosophila melanogaster.  
XX  
PN WO200171042-A2.  
PD 27-SEP-2001.  
XX  
PF 23-MAR-2001; 2001WO-US09231.  
XX  
PR 23-MAR-2000; 2000US-191637P.  
PR 11-JUL-2000; 2000US-0614150.  
XX  
PA (PEKE ) PE CORP NY.  
XX  
PI Venter JC, Adams M, Li PWD, Myers EW;  
XX  
DR WPI; 2001-656860/75.  
DR N-PSDB; ABL07077.  
XX  
PT New isolated nucleic acid detection reagent for detecting 1000 or more  
PT genes from Drosophila and for elucidating cell signalling and cell-cell  
PT interactions -  
XX  
PS Disclosure; SEQ ID NO 15714; 21pp + Sequence Listing; English.  
XX

The invention relates to an isolated nucleic acid detection reagent  
CC capable of detecting 1000 or more genes from Drosophila. The invention is  
CC useful in developmental biology and in elucidating cell signalling and  
CC cell-cell interactions in higher eukaryotes for the development of  
CC insecticides, therapeutics and pharmaceutical drugs. The invention  
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA  
CC sequences (ABL01840-ABL16175) and the encoded proteins  
CC (ABB57737-ABB72072).  
CC  
CC The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 202 AA;

Query Match 9.0%; Score 103; DB 22; Length 202;  
Best Local Similarity 28.3%; Pred. No. 0.1;  
Matches 34; Conservative 12; Mismatches 48; Indels 26; Gaps 4;

QY 60 GGGGGGVGPFRRWGTGLRMRCQQQLQDVSRRCRCAAIRSMVRGYEAMPPL 114  
Db 72 GGGGGGGVGGVQGGYQGPAPLTEQVGYGGTQ---GGYGSAAKGYGASVPPEFALGLSGQG 128

QY 115 PWGRQQQPPPGG---GGGGGGYYPGSRPCEGYGYGGGQRMYPPCRPGTTGGGPRI 170  
Db 129 RLGGYGNATPQSGKLGYGGGGGYRRPAPQSQLYAAGE-----LGAGPQL 174

Search completed: November 29, 2003, 17:05:01  
Job time : 62 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 29, 2003, 16:58:04 ; Search time 26 Seconds  
(without alignments)  
761.952 Million cell updates/sec

Title: US-10-053-410-4  
Perfect score: 1148  
Sequence: 1 MAKIAAAAAAALCFALVAV.....MMCRLEPQECISFGSDQY 206  
Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96169682 residues  
Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR 76:\*  
1: pir1:\*  
2: pir2:\*  
3: pir3:\*  
4: pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	439	38.2	186	JC4784	alpha-globulin pre
2	435	37.9	186	WMR219	19K globulin precu
3	289.5	25.2	660	A24266	glutenin high mole
4	287	25.0	705	S18733	glutenin high mole
5	282.5	24.6	707	S52390	D-hordein precu
6	281	24.5	648	S04832	glutenin high mole
7	157	13.7	830	S15720	glutenin high mole
8	156	13.6	791	JN0690	glutenin, high-mol
9	156	13.6	815	S03843	glutenin high mole
10	155.5	13.5	848	S02262	glutenin high mole
11	152	13.2	815	JN0689	glutenin high mole
12	151.5	13.2	789	A30843	glutenin, high-mol
13	151.5	13.2	838	1 EEWTHW	glutenin, high mol
14	147.5	12.8	733	JC2099	glutenin, high mol
15	133.5	11.6	161	2 S38887	high-molecular-wei
16	126	11.0	323	S38887	2S albumin - commo
17	117.5	10.2	309	2 S10889	proline-rich prote
18	114.5	10.0	146	S14946	2S seed storage pr
19	114	9.9	297	2 T06500	alpha/beta-gliadin
20	112.5	9.8	258	1 RZCS	2S seed storage pr
21	112	9.8	295	2 S01062	2S seed storage pr
22	111.5	9.7	307	2 S10015	alpha/beta-gliadin
23	111	9.7	296	2 S07361	alpha/beta-gliadin
24	110	9.6	101	1 EEW1	glutenin 1 - wheat
25	110	9.6	139	2 T09878	albumin 2S storage
26	110	9.6	145	2 S13376	CM2 protein - duru
27	110	9.6	326	2 A41732	heterogeneous ribo
28	110	9.6	386	1 S22315	snRNP-associated p
29	109.5	9.5	291	2 T06498	alpha/beta-gliadin

30	109	9.5	160	2 S59923	allergen RA14C pre
31	109	9.5	166	2 S31082	seed allergen RAG2
32	109	9.5	319	2 C22364	alpha/beta-gliadin
33	108.5	9.5	154	2 S14947	2S albumin - Bzai
34	108.5	9.5	280	2 G84839	late embryogenesis
35	108	9.4	220	2 A44805	eggshell protein p
36	108	9.4	806	2 T13690	hypothetical prote
37	106.5	9.3	145	2 S10027	alpha-amylase inhi
38	106.5	9.3	313	2 S07924	alpha/beta-gliadin
39	106	9.2	212	2 B36298	proline-rich prote
40	105.5	9.2	141	2 T10257	2S albumin precurs
41	105	9.1	135	1 T12M1	trypsin/factor XII
42	105	9.1	1733	1 A45344	probable nuclear a
43	104.5	9.1	220	2 A36298	proline-rich prote
44	104	9.1	139	2 T09850	albumin 2S storage
45	104	9.1	148	1 T1BH	trypsin inhibitor

ALIGNMENTS

RESULT 1

JC4784  
alpha-globulin precursor - rice  
C:Species: Oryza sativa (rice)  
C:Date: 10-May-1996 #sequence\_revision 16-Aug-1996 #text\_change 20-Jun-2000  
C:Accession: JC4784  
R:Nakase, M.; Hotta, H.; Adachi, T.; Aoki, N.; Nakamura, R.; Masumura, T.; Tanaka, K.; Me  
Gene 170, 223-226, 1996  
A:Title: Cloning of the rice seed alpha-globulin-encoding gene: Sequence similarity of t  
A:Reference number: JC4784; MUID:96235139; PMID:8666249  
A:Accession: JC4784  
A:Molecule type: DNA  
A:Residues: 1-186 <NAK>  
A:Cross-references: DDBJ:P50643; NID:G840704; PIDN:BAA09308.1; PID:G1783206  
A:Experimental source: seed  
C:Genetics:  
A:Gene: Glb  
C:Superfamily: wheat alpha-amylase inhibitor  
C:Keywords: Globulin; seed  
F:1-23/Domain: signal sequence #status predicted <SIG>  
F:24-186/Product: alpha-globulin #status predicted <MAT>

Query Match	38.2%	Score 439;	DB 2;	Length 186;
Best Local Similarity	51.4%	Pred. No. 1.6e-29;		
Matches 108;	Conservative 20;	Mismatches 54;	Indels 28;	Gaps 9;
QY	1	MAKIAAAAAAALCFALVAVAVCQGEVERQRLRLDQCWQEVQESPLDACRQVLDRQLTGG	60	
DB	1	MASKVVFFAAL-MAAMVAISGAQLSESEMRFDRCQREVQDSPLDACRQVLDRQLTGR	59	
QY	61	GGGGVGPRFWGTGLRMRCQQLQDVSRRCRAATRSVMRGYEAMP-PLEKGWMPWGRQ	119	
DB	60	ERFQPMFRFPFGALGLRMQCQQLQDVSRRCRAAIRMVRSYEESMPLEQGWSSSSE	119	
QY	120	QPPPGGGGGGGYYPGCRPGEGYGGQGORQMY---PPCRPGTTGGPPRIGVRILT	176	
DB	120	YY-----GGEG-----SSSEQGY-YGEGSEEGYGEQQQPGMT-----RVRLT	158	
QY	177	KAREYAAGLPMMCRLEPQECISFGSDQY	206	
DB	159	RARQYAAQLPMCRV-EPOQCSIFAAG-QY	186	

RESULT 2

WMR219  
19K globulin precursor - rice  
N:Alternate names: alpha-globulin  
C:Species: Oryza sativa (rice)  
C:Date: 31-Dec-1992 #sequence\_revision 31-Dec-1992 #text\_change 18-Jun-1999  
C:Accession: S20024; S25735; PN0497  
R:Shorrosh, B.S.; Wen, L.; Zen, K.C.; Huang, J.K.; Pan, J.S.; Hermodson, M.A.; Tanaka, K.

Plant Mol. Biol. 18, 151-154, 1992  
A;Title: A novel cereal storage protein: molecular genetics of the 19 kDa globulin of rice  
A;Reference number: S20024; MUID:92119226; PMID:11731968  
A;Accession: S20024  
A;Molecule type: mRNA  
A;Residues: 1-186 <SHO>  
A;Cross-references: EMBL:X63990; NID:g20158; PIDN:CAA45400.1; PID:g20159  
A;Accession: S25735  
A;Molecule type: protein  
A;Residues: 66-74;108-133;171-186 <SH2>  
R;Krishnan, H.B.; Pueppke, S.G.  
Biochem. Biophys. Res. Commun. 193, 460-466, 1993  
A;Title: Nucleotide sequence of an abundant rice seed globulin: homology with the high molecular weight chain 12 precursor - wheat  
A;Reference number: PN0497; MUID:93277591; PMID:8503935  
A;Accession: PN0497  
A;Molecule type: mRNA  
A;Residues: 6-186 <KRI>  
A;Cross-references: GB:L12252  
A;Experimental source: seed  
A;Superfamily: wheat alpha-amylase inhibitor  
C;Keywords: storage protein  
F;1-2/Domain: signal sequence #status predicted <SIG>  
F;23-186/Product: 19K globulin #status predicted <MAT>  
  
Query Match 37.9%; Score 435; DB 1; Length 186;  
Best Local Similarity 51.0%; Pred. No. 3.4e-29;  
Matches 107; Conservative 20; Mismatches 55; Indels 28; Gaps 9;  
  
QY 1 MAKIAAAAAALCFALVAVAVCQGEVERQRLDLQCMQVQESPLDACRQVLDRLTGG 60  
DB 1 MAAKLVLFVAVV--TALVALTTAEASRQ-----LQCERELQESSLEACRQVVDQQL--- 51  
QY 61 GGGGGVGFPRFWGTGLRMRCQQQLQDVSRRCRAIRSMVRGYEE-AMPLEKGMWP---- 115  
DB 52 -----AGRLPWTSTGLQMRCCQQLRDVSAKCRSVAVSQVARYEQTVVPPKGGSFYFGETT 106  
QY 116 -----WG-----RQ-----QPPPPQGGGGGGG----- 133  
DB 107 PLQQLQQGIFWGTSSQTQGYPSVTSPRQSGYVPGQASFPQGGQQPGKMQEPGQGO 166  
QY 134 YYPGCS--RPGGGYGVGGQGRQMYPPCRPGTTGGPRIGR 172  
DB 167 WYPTSLQPPGGGQQIGKG--KQGYPTSLQPPGGGQQIGQ 205  
  
RESULT 4  
S18733  
glutenin high molecular weight chain 18y9 precursor - wheat  
C;Species: Triticum aestivum (common wheat)  
C;Date: 08-Jun-1994 #sequence\_revision 10-Nov-1995 #text\_change 20-Aug-1999  
C;Accession: S18733  
R;Halford, N.G.; Forde, J.; Anderson, O.D.; Greene, F.C.; Shewry, P.R.  
Theor. Appl. Genet. 75, 117-126, 1987  
A;Title: The nucleotide and deduced amino acid sequences of an HMW glutenin subunit gene A and 1D.  
A;Reference number: S18733  
A;Accession: S18733  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-705 <HAL>  
A;Cross-references: EMBL:X61026; NID:g22089; PIDN:CAA43361.1; PID:g22090  
C;Superfamily: glutenin  
  
Query Match 25.0%; Score 287; DB 2; Length 705;  
Best Local Similarity 32.9%; Pred. No. 2.3e-16;  
Matches 85; Conservative 24; Mismatches 47; Indels 102; Gaps 13;  
  
QY 17 LVAVAVCQGEVERQRLDLQCMQVQESPLDACRQVLDRLTGGGGGGVGFPRFWGTGLR 76  
DB 15 LVALTAAGEASRQ---LQCERELQESSLEACRQVVDQQL-----AGRLPWTSTGLQ 62  
QY 77 MRCCQQLQDVSRRCRAIRSMVRGYEE-AMPLEKGMWP-----WGR--- 118  
DB 63 MRCCQQLRDVSAKCRPVAVSQVRYEQTVVPPKGGSFYFGETTPLQQLQQVIFWGTSSQ 122  
QY 119 -----QQQP-----PQGG-----GGGGG----- 134  
DB 123 TVQGYPSVSSPQQGYPYPGQASFPQGGQQPGKMQELGCGQQGYPTSLHSGGGGQGO 182  
QY 135 YYPGCS--RPGEGYGVGG-----GQRMYP--PCRPG---TTGGG 167  
DB 183 YFPSSLLQPPGGGQQIGGQGGYPTSLQPPGGGQQIGGQGGYPTSPQHPGQRQPGGQ 242  
QY 168 PRIGR-VRLTKAREYAAAG 184  
DB 243 QQIGGGQQLGQGRQIGQ 260  
  
RESULT 5  
S52390  
D-hordein precursor - barley  
C;Species: Hordeum vulgare (barley)  
C;Date: 08-May-1995 #sequence\_revision 26-Jul-1996 #text\_change 20-Aug-1999  
C;Accession: S52390; S23921  
R;Sorensen, M.B.; Muller, M.; Simpson, D.  
submitted to the EMBL Data Library, February 1995  
A;Description: Hordein promoter methylation and transcriptional activity in wild type and  
A;Reference number: S52390  
A;Accession: S52390  
A;Molecule type: DNA

Plant Mol. Biol. 18, 151-154, 1992  
A;Title: A novel cereal storage protein: molecular genetics of the 19 kDa globulin of rice  
A;Reference number: S20024; MUID:92119226; PMID:11731968  
A;Accession: S20024  
A;Molecule type: mRNA  
A;Residues: 1-186 <SHO>  
A;Cross-references: EMBL:X63990; NID:g20158; PIDN:CAA45400.1; PID:g20159  
A;Accession: S25735  
A;Molecule type: protein  
A;Residues: 66-74;108-133;171-186 <SH2>  
R;Krishnan, H.B.; Pueppke, S.G.  
Biochem. Biophys. Res. Commun. 193, 460-466, 1993  
A;Title: Nucleotide sequence of an abundant rice seed globulin: homology with the high molecular weight chain 12 precursor - wheat  
A;Reference number: PN0497; MUID:93277591; PMID:8503935  
A;Accession: PN0497  
A;Molecule type: mRNA  
A;Residues: 6-186 <KRI>  
A;Cross-references: GB:L12252  
A;Experimental source: seed  
A;Superfamily: wheat alpha-amylase inhibitor  
C;Keywords: storage protein  
F;1-2/Domain: signal sequence #status predicted <SIG>  
F;23-186/Product: 19K globulin #status predicted <MAT>  
  
Query Match 37.9%; Score 435; DB 1; Length 186;  
Best Local Similarity 51.0%; Pred. No. 3.4e-29;  
Matches 107; Conservative 20; Mismatches 55; Indels 28; Gaps 9;  
  
QY 1 MAKIAAAAAALCFALVAVAVCQGEVERQRLDLQCMQVQESPLDACRQVLDRLTGG 60  
DB 1 MAAKLVLFVAVV--TALVALTTAEASRQ-----LQCERELQESSLEACRQVVDQQL--- 51  
QY 61 GGGGGVGFPRFWGTGLRMRCQQQLQDVSRRCRAIRSMVRGYEE-AMPLEKGMWP---- 115  
DB 52 -----AGRLPWTSTGLQMRCCQQLRDVSAKCRSVAVSQVARYEQTVVPPKGGSFYFGETT 106  
QY 116 -----WG-----RQ-----QPPPPQGGGGGGG----- 133  
DB 107 PLQQLQQGIFWGTSSQTQGYPSVTSPRQSGYVPGQASFPQGGQQPGKMQEPGQGO 166  
QY 134 YYPGCS--RPGGGYGVGGQGRQMYPPCRPGTTGGPRIGR 172  
DB 167 WYPTSLQPPGGGQQIGKG--KQGYPTSLQPPGGGQQIGQ 205  
  
RESULT 4  
S18733  
glutenin high molecular weight chain 12 precursor - wheat  
C;Species: Triticum aestivum (common wheat)  
C;Date: 09-Sep-1987 #sequence\_revision 09-Sep-1987 #text\_change 20-Aug-1999  
C;Accession: A24266; S06644  
R;Thompson, R.D.; Bartels, D.; Harberd, N.P.  
Nucleic Acids Res. 13, 6833-6846, 1985  
A;Title: Nucleotide sequence of a gene from chromosome 1D of wheat encoding a HMW-glutenin subunit  
A;Reference number: A24266; MUID:96041882; PMID:3840588  
A;Accession: A24266  
A;Molecule type: DNA  
A;Residues: 1-660 <THO>  
A;Cross-references: GB:X03041; NID:g21778; PIDN:CAA26847.1; PID:g21779  
A;Experimental source: cv. Chinese Spring  
R;Goldsbrough, A.P.; Bulleld, N.J.; Freedman, R.B.; Flavell, R.B.  
Biochem. J. 263, 837-842, 1989  
A;Title: Conformational differences between two wheat (Triticum aestivum) 'high-molecular weight' glutenin subunits  
A;Reference number: S06644; MUID:90088430; PMID:2597130  
A;Accession: S06644  
A;Status: not compared with conceptual translation  
A;Molecule type: DNA  
A;Residues: 126-660 <GOL>  
C;Superfamily: glutenin  
  
Query Match 25.2%; Score 289.5; DB 2; Length 660;  
Best Local Similarity 36.2%; Pred. No. 1.4e-16;  
Matches 80; Conservative 24; Mismatches 52; Indels 65; Gaps 10;

A;Residues: 1-475 <SOR>  
A;Cross-references: EMBL:X84368; NID:g671536; PIDN:CAA59104.1; PID:g671537  
R;Hallford, N.G.; Tatham, A.S.; Sui, E.; Daxoda, L.; Dreyer, T.; Shewry, P.R.  
Biochim. Biophys. Acta 1122, 118-122, 1992  
A;Title: Identification of a novel beta-turn-rich repeat motif in the D hordeins of barley  
A;Reference number: S23921; MUID:92353095; PMID:1643086  
A;Accession: S23921  
A;Molecule type: mRNA  
A;Residues: 267-355, 'P', 357-359, 'Q', 361-458, 'Y', 460-707 <HAL>  
A;Cross-references: EMBL:X68072; NID:g18969; PIDN:CAA48209.1; PID:g18970  
C;Genetics:  
A;Gene: hor3  
C;Superfamily: glutenin  
C;Keywords: seed; storage protein  
F;1-21/Domain: signal sequence #status predicted <SIG>  
F;22-707/Product: D-hordein #status predicted <MAT>

Query Match 24.6%; Score 282.5; DB 2; Length 707;  
Best Local Similarity 36.2%; Pred. No. 5.5e-16;  
Matches 79; Conservative 22; Mismatches 52; Indels 65; Gaps 11;

QY 1 MAKIAAAAAALCFALVAVAVCOGEVE-----RORLRDQCWQVEQSPDLACRQVLD 54  
Db 1 MAKRLVLFVAVI--VALVALTTAERINGNIFLDSRQLQCERELQSSLEACRRVD 58  
QY 55 RQLTGGGGGGVGFRTGLMRCCQQLQDVSRRCRAIRSMVRGYEE----- 104  
Db 59 QQL-----VGQLPWTSTGLQWCCQLQDLVDSPCRPVALSQVRYEQTEVPKSGGS 110  
QY 105 -----AMPLEK-GWV-----PWGRQQPPPGGGGGGGGGY-----YPC 138  
Db 111 FYPGGTAPPLQGGWGTSTKNNYPDTSSQSQWQGGYHQSVTSQQPGQGGQSYFG 170  
QY 139 S-----RPGGYGVGCGGQRO--MYP-----PCRGTGG 166  
Db 171 STFPQPGGQ-----QPGQRQPSYATPFPQPGGQGG 205

RESULT 6  
S04832  
glutenin high molecular weight chain (Dy10) - wheat  
C;Species: Triticum aestivum (common wheat)  
C;Date: 28-Feb-1990 #sequence\_revision 28-Feb-1990 #text\_change 20-Aug-1999  
C;Accession: S04832; S06645  
R;Anderson, O.D.; Greene, F.C.; Yip, R.E.; Halford, N.G.; Shewry, P.R.; Malpica-Romero,  
Nucleic Acids Res. 17, 461-462, 1989  
A;Title: Nucleotide sequences of the two high-molecular-weight glutenin genes from the D  
A;Reference number: S02262; MUID:89098419; PMID:2563152  
A;Accession: S04832  
A;Status: not compared with conceptual translation  
A;Molecule type: DNA  
A;Residues: 1-648 <AND>  
A;Cross-references: GB:X12929; NID:gl360617; PIDN:CAA31396.1; PID:g21751  
R;Goldsbrough, A.P.; Bulleid, N.J.; Freedman, R.B.; Flavell, R.B.  
Biochem. J. 263, 837-842, 1989  
A;Title: Conformational differences between two wheat (Triticum aestivum) 'high-molecular  
A;Reference number: S06644; MUID:90088430; PMID:2597130  
A;Accession: S06645  
A;Status: not compared with conceptual translation  
A;Molecule type: DNA  
A;Residues: 126-474, 'QG', 477-648 <GOL>  
C;Genetics:  
A;Gene: Glu-D1-2b  
C;Superfamily: glutenin

Query Match 24.5%; Score 281; DB 2; Length 648;  
Best Local Similarity 36.6%; Pred. No. 6.8e-16;  
Matches 78; Conservative 23; Mismatches 46; Indels 66; Gaps 11;

QY 1 MAKIAAAAAALCFALVAVAVCOGEVEQRLRDLQCWQVEQSPDLACRQVLDRLTGG 60  
Db 1 MAKRLVLFVAVV--IALVALTTAEGEASQ-----LQCERELQSSLEACRRQVVDQQL--- 51

QY 61 GGGGGVGFRTGLMRCCQQLQDVSRRCRAIRSMVRGYEE-AMPLEKGWMP----- 115  
Db 52 -----AGRLPWTSTGLMRCCQQLRVDVSAKRSVAVSQARQYEQTVVPPKGGSYFGETT 106  
QY 116 -----WG-----RQ-----QPPPGGGGGGGG----- 133  
Db 107 PLQQLQGGIFWGTSTQTVGYYPGVTSRQSGSYFPGQASPPQPGQCGKQWEPQGGQ 166  
QY 134 YYPGCS--RPGEGYGVGCGGQRMVPPC--RPG 162  
Db 167 WYFPTSLQPGGGGQGGQIGK-GQGGYFPTSLOQPG 198

RESULT 7  
S15720  
glutenin high molecular weight chain 1A1 precursor - wheat  
C;Species: Triticum aestivum (common wheat)  
C;Date: 08-Jun-1994 #sequence\_revision 12-May-1995 #text\_change 20-Aug-1999  
C;Accession: S15720  
R;Hallford, N.G.; Field, J.M.; Blair, H.; Urwin, P.; Moore, K.; Robert, L.; Thompson, R.;  
submitted to the EMBL Data Library, July 1991  
A;Description: Analysis of HMW glutenin subunit encoded by chromosome 1A of bread wheat:  
A;Reference number: S15720  
A;Accession: S15720  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-830 <HAL>  
A;Cross-references: EMBL:X61009; NID:g21742; PIDN:CAA43331.1; PID:g21743  
C;Superfamily: glutenin

Query Match 13.7%; Score 157; DB 2; Length 830;  
Best Local Similarity 24.6%; Pred. No. 1.7e-05;  
Matches 63; Conservative 17; Mismatches 42; Indels 134; Gaps 11;

QY 11 ALCFALVAVAVCOGEVEQRLRDLQCWQVEQSPDLACRQVLDRLTGGGGGGVGPFR 70  
Db 9 AAVVALVALTAEGEASQ-----LQCERELQSSLEKACRQVVD----- 48  
QY 71 WGTGLMRCCQQLQDVSRRCRAIRSMVRGYEE-AMPLEKGWMP----- 115  
Db 49 -----QQLRVDVSPQCQVGGGFVARQYEQVVPKGGSYFGETTTPQQLQSSI 98  
QY 116 -W-----GQQQP-----PQGG----- 127  
Db 99 LAGIPALLRRYLVTSPPQSVYFPCQASSRPGGQGGQGGQYVLTSPQSGQWQPP 158  
QY 128 GGGGGY-----YPPCS-----RPGGYGVGQ----- 150  
Db 159 GGGQAGYPTSPQSGQGGQYFYTSPWQPEQLQPTQGGQRLRQGGQGGQS 218  
QY 151 --GQRMVPPC--RPG 162  
Db 219 GGGQPRYPTSPQPG 234

RESULT 8  
JN0690  
glutenin, high-molecular-weight Bx7 chain precursor - wheat  
C;Species: Triticum aestivum (common wheat)  
C;Date: 14-Jul-1994 #sequence\_revision 14-Jul-1994 #text\_change 14-Jul-1994  
C;Accession: JN0690  
R;Anderson, O.D.; Greene, F.C.  
Theor. Appl. Genet. 77, 689-700, 1989  
A;Title: The characterization and comparative analysis of high-molecular-weight glutenin  
A;Reference number: JN0689  
A;Accession: JN0690  
A;Molecule type: DNA  
A;Residues: 1-791 <AND>  
C;Comment: The main wheat storage proteins are divided into two groups. The glutenins, c  
A families.  
C;Superfamily: glutenin  
C;Keywords: seed; storage protein  
F;1-21/Domain: signal sequence #status predicted <SIG>

F;22-791/Product: glutenin, HMW Bx7 chain #status predicted <MAT>

Query Match 13.6%; Score 156; DB 2; Length 791;  
Best Local Similarity 28.1%; Pred. No. 2e-05;  
Matches 59; Conservative 19; Mismatches 54; Indels 78; Gaps 10;  
QY 1 MAKIAAAAAALCFALVAVAVCQGEVERQLRDLCQWQVQESPLDACRQVLDRLQTGG 60  
DB 1 MAKRLVLFVAVV--VALVALTAAGEASGQ---LQC-----EHELEACQVVD----- 43  
QY 61 GGGGGVGPFRWGTGLRMRCQQQLQDVSRRCRAIRSMVRGYEE--AMPPEKXGWWP--- 115  
DB 44 -----QQQLRDVSPGCRPTVSPGTRQYEQQPVWFSKAGSFYPSSET 83  
QY 116 -----WG-----RQQP-----PPQGGGGGGGGYYPYPCSRPGE 143  
DB 84 TPSQQLQOMIFWGIIPALLRRYPSVTSSQGSYYPGQSPQSGGQGGQPGQE---QQPGQ 140  
QY 144 GYGYGGGGGROM-YPCRCPTGGPRIGR 172  
DB 141 GQHQQPGQRQQYYTSPQPGQGGQQLGQ 170

RESULT 9

B30843

glutenin high molecular weight chain Ax2 precursor - wheat

C;Species: Triticum aestivum (common wheat)  
C;Date: 19-May-1989 #sequence\_revision 19-May-1989 #text\_change 15-Nov-1996  
C;Accession: B30843  
R;Anderson, O.D.; Greene, F.C.  
submitted to GenBank, January 1989  
A;Reference number: A94515  
A;Accession: B30843  
A;Molecule type: DNA  
A;Residues: 1-815 <AND>  
C;Superfamily: Glutenin  
C;Keywords: seed; storage protein  
F;1-21/Domain: signal sequence #status predicted <SIG>  
F;22-815/Product: glutenin Ax2 chain #status predicted <MAT>  
F;108-773/Region: glutamine/glycine/proline-rich

Query Match 13.6%; Score 156; DB 2; Length 815;  
Best Local Similarity 24.8%; Pred. No. 2.1e-05;  
Matches 62; Conservative 17; Mismatches 43; Indels 128; Gaps 11;  
QY 11 ALCFALVAVAVCQGEVERQLRDLCQWQVQESPLDACRQVLDRLQTGGGGGGVGPFR 70  
DB 9 AAVVALVALTAAGEASGQ---LQCERELQEHSLKACRQVVD----- 48  
QY 71 WGTGLRMRCQQQLQDVSRRCRAIRSMVRGYEE--AMPPEKXGWWP----- 115  
DB 49 -----QQQLRDVSPGCRPTVSPGTRQYEQQPVWFSKAGSFYPSSET 98  
QY 116 -WG-----RQQP-----PPQGG-----GGGGGG 133  
DB 99 LMGIPALLRRYLSVTSPPQVSYYPGQASSRQRCQGEVYLTSPQSGQWQPGGQSG 158  
QY 134 Y-----YPPCS-----RPGEGYGYGG-----GQRQ 154  
DB 159 YPTSPQSGQKPGYYPYTPWQPELQQTQGGQRRQPGQQLRQGGQGGQGGQGP 218  
QY 155 MYPPC--RPG 162  
DB 219 YPTSSQQPG 228

RESULT 10

S02262

glutenin high molecular weight chain Dx5 - wheat

C;Species: Triticum aestivum (common wheat)  
C;Date: 28-Feb-1990 #sequence\_revision 28-Feb-1990 #text\_change 03-Feb-1994  
C;Accession: S02262  
R;Anderson, O.D.; Greene, F.C.; Yip, R.E.; Halford, N.G.; Shewry, P.R.; Malpica-Romero,

Nucleic Acids Res. 17, 461-462, 1989

A;Title: Nucleotide sequences of the two high-molecular-weight glutenin genes from the D.  
A;Reference number: S02262, MUID:89098419; PMID:2563152  
A;Accession: S02262  
A;Molecule type: DNA  
A;Residues: 1-848 <AND>  
A;Cross-references: EMBL:X12928  
C;Genetics:  
A;Gene: Glu-D1-1b  
C;Superfamily: glutenin

Query Match 13.5%; Score 155.5; DB 2; Length 848;  
Best Local Similarity 26.0%; Pred. No. 2.3e-05;  
Matches 61; Conservative 17; Mismatches 52; Indels 105; Gaps 10;  
QY 1 MAKIAAAAAALCFALVAVAVCQGEVERQLRDLCQWQVQESPLDACRQVLDRLQ 57  
DB 1 MAKRLVLFVAVV--VALVALTAAGEASGQ---LQCERELQEHSLKACQVVD--- 51  
QY 58 TGGGGGGVGPFRWGTGLRMRCQQQLQDVSRRCRAIRSMVRGYEE--AMPPEKXGWWP 115  
DB 52 -----QQQLRDVSPGCRPTVSPGTRQYEQQPVWFSKAGSFYPSSET 89  
QY 116 -----WG-----RQQP-----GGGGGGG 134  
DB 89 GETTPPQQLQORIFWGIIPALLRRYPSVTCPOQVSYYPGQASSRQRCQGEVYLTSPQSGQWQPGGQGY 148  
QY 135 -----YPPCS-----RPGEGYGYGGGGQRCQWYPPCRPG 162  
DB 149 PTPSPQPGQWQPGQEQPRYPTSPQSGQLQPAQGGQGGQGGQGGQGGQPG 203

RESULT 11

JN0689

glutenin, high-molecular-weight Ax2\* chain precursor - wheat

C;Species: Triticum aestivum (common wheat)  
C;Date: 14-Jul-1994 #sequence\_revision 14-Jul-1994 #text\_change 15-Nov-1996  
C;Accession: JN0689  
R;Anderson, O.D.; Greene, F.C.  
Theor. Appl. Genet. 77, 689-700, 1989  
A;Title: The characterization and comparative analysis of high-molecular-weight glutenin  
A;Reference number: JN0689  
A;Accession: JN0689  
A;Molecule type: DNA  
A;Residues: 1-815 <AND>  
A;Note: The authors translated the codon CTA for residue 11 as Val, CAT for residue 496 &  
C;Comment: The main wheat storage proteins are divided into two groups. The glutenins, c  
a families.  
C;Genetics:  
A;Gene: Ax2\*

C;Superfamily: glutenin

F;1-21/Domain: signal sequence #status predicted <SIG>

F;22-815/Product: glutenin, high-molecular-weight Ax2\* chain #status predicted <MAT>

Query Match 13.2%; Score 152; DB 2; Length 815;  
Best Local Similarity 24.4%; Pred. No. 4.4e-05;  
Matches 61; Conservative 18; Mismatches 43; Indels 128; Gaps 11;  
QY 11 ALCFALVAVAVCQGEVERQLRDLCQWQVQESPLDACRQVLDRLQTGGGGGGVGPFR 70  
DB 9 AALVALVALTAAGEASGQ---LQCERELQEHSLKACQVVD----- 48  
QY 71 WGTGLRMRCQQQLQDVSRRCRAIRSMVRGYEE--AMPPEKXGWWP----- 115  
DB 49 -----QQQLRDVSPGCRPTVSPGTRQYEQQPVWFSKAGSFYPSSET 98  
QY 116 -WG-----RQQP-----PPQGG-----GGGGGG 133  
DB 99 LMGIPALLRRYLSVTSPPQVSYYPGQASSRQRCQGEVYLTSPQSGQWQPGGQSG 158  
QY 134 Y-----YPPCS-----RPGEGYGYGG-----GQRQ 154  
DB 159 YPTSPQSGQKPGYYPYTPWQPELQQTQGGQRRQPGQQLRQGGQGGQGGQGP 218



QY 155 MYPPC--RPG 162  
Db 219 YNPTSSQPG 228

RESULT 12  
A30843  
glutenin high molecular weight chain Bx7 precursor - wheat  
C:Species: Triticum aestivum (common wheat)  
C:Date: 19-May-1989 #sequence\_revision 19-May-1989 #text\_change 20-Aug-1999  
C:Accession: A30843  
R:Anderson, O.D.; Greene, F.C.;  
submitted to GenBank, January 1989  
A:Reference number: A94515  
A:Accession: A30843  
A:Molecule type: DNA  
A:Residues: 1-789 <AND>  
A:Cross-references: EMBL:M22209; NID:g170744; PIDN:AAA34291.1; PID:g170745  
C:Superfamily: glutenin  
C:Keywords: seed; storage protein  
F:1-21/Domain: signal sequence #status predicted <SIG>  
F:22-789/Product: glutenin Bx7 chain #status predicted <MAT>  
F:103-747/Region: glutamine/glycine/proline-rich

Query Match 13.2%; Score 151.5; DB 2; Length 789;  
Best Local Similarity 24.4%; Pred. No. 4.7e-05;  
Matches 62; Conservative 21; Mismatches 52; Indels 119; Gaps 9;

QY 1 MAKIAAAAAALCFALVAVAVCOGEVERQRLDLQCWQEVQESPLDACRQVLDRLTGG 60  
Db 1 MAKRLVLFVAVV--VALVALTAAGEASGQ-----LQC-----EHELEACQVVD----- 43

QY 61 GGGGGVGFPRWGTLRMRCQQLQDVSRRCRAAIRSMVRGYEE----- 104  
Db 44 -----QQLRDVSPGCRPITVSPGTRQYEQQPVVASKAGSFYPSPT 83

QY 105 -----AMPPLEKGNWPW-----GROQPP----- 122  
Db 84 TFSQQLQQLFWGIPALLRRYPSVTSSQGSYYPGQASPGSQGQPGQEQPGQGGQ 143

QY 123 -----PPQGG-----GGGGGYYPCSRPGEVGYGQGGQ-----RQMYPP 158  
Db 144 DQPGGQGGYYPTSPQPGQGGQGGQGGYYPTSPQPGQKQAGGQGGQGGQGGYY 203

QY 159 CRPGTGGGPRIGR 172  
Db 204 TSPQSGGQGGQPGQ 217

RESULT 13  
EETHW  
glutenin, high molecular weight chain precursor - wheat  
C:Species: Triticum aestivum (common wheat)  
C:Date: 30-Jun-1988 #sequence\_revision 30-Jun-1988 #text\_change 16-Jul-1999  
C:Accession: A24107  
R:Sugiyama, T.; Rafalski, A.; Peterson, D.; Soll, D.  
Nucleic Acids Res. 13, 8729-8737, 1985.  
A:Title: A wheat HMW glutenin subunit gene reveals a highly repeated structure.  
A:Reference number: A24107; MUID:86093674; PMID:3001648  
A:Accession: A24107  
A:Molecule type: DNA  
A:Residues: 1-838 <SUG>  
A:Cross-references: GB:X03346; NID:g21784; PIDN:CAA27052.1; PID:g736319  
A:Experimental source: cv. Yanhill  
C:Comment: Glutensins, like gliadins, are high in glutamine and proline but differ in con  
C:Superfamily: glutenin  
C:Keywords: seed; storage protein  
F:1-21/Domain: signal sequence #status predicted <SIG>  
F:22-838/Product: glutenin, HMW chain #status predicted <MAT>

Query Match 13.2%; Score 151.5; DB 1; Length 838;  
Best Local Similarity 28.7%; Pred. No. 5e-05;

Matches 58; Conservative 18; Mismatches 45; Indels 81; Gaps 10;

QY 1 MAKIAAAAAALCFALVAVAVCOGEVERQRLDLQCWQEVQESPLDACRQVLDRLQ 57  
Db 1 MAKRLVLFVAVV--VALVALTAAGEASGQ-----LQCERELQELQERELXACQVMD--- 51

QY 58 TGGGGGGVGFPRWGTLRMRCQQLQDVSRRCRAAIRSMVRGYEE----- 104  
Db 52 -----QQLRDVSPGCRPITVSPGTRQYEQQPVVASKAGSFYPSPT 88

QY 105 -----AMPPLEKGNWPW-----GROQPP----- 122  
Db 89 ETTTPPQQLQORIFWGIPALLRRYPSVTSPQGSYYPGQASPGSQGQGGQGGYY 142

QY 143 EGVGYGQGGQGGQGGQGGQGGQGGQGGQGGQGGQGGQGGQGGQGGQGGQGGQ 140  
Db 141 QSQSGQ--GQGGYYPTSPQPG 161

## RESULT 14

JC2099

glutenin, high molecular weight chain Bx17 - wheat

C:Species: Triticum aestivum (common wheat)

C:Date: 14-Jul-1994 #sequence\_revision 14-Jul-1994 #text\_change 23-Mar-1995

C:Accession: JC2099; JQ2261

R:Reddy, P.; Appels, R.

Theor. Appl. Genet. 85, 616-624, 1993

A:Title: Analysis of a genomic DNA segment carrying the wheat high-molecular-weight

A:Reference number: JC2099

A:Accession: JC2099

A:Molecule type: DNA

A:Residues: 1-753 &lt;RED&gt;

A:Experimental source: leaf

C:Genetics:

A:Gene: Bx17

A:Map position: 1B

C:Superfamily: glutenin

C:Keywords: seed; storage protein

Query Match 12.8%; Score 147.5; DB 2; Length 753;  
Best Local Similarity 23.8%; Pred. No. 9.7e-05;  
Matches 58; Conservative 20; Mismatches 49; Indels 117; Gaps 8;

QY 11 ALCFALVAVAVCOGEVERQRLDLQCWQEVQESPLDACRQVLDRLTGGGGGGVGPFR 70  
Db 9 AAVVVALVALTAAGEASGQ-----LQC-----EHELEACQVVD----- 43

QY 71 WGTGLMRCCQQLQDVSRRCRAAIRSMVRGYEE----- 104  
Db 44 -----QQLRDVSPGCRPITVSPGTRQYEQQPVVASKAGSFYPSPT 93

QY 105 -----AMPPLEKGNWPW-----GROQPP----- 122  
Db 94 FWGIPALLRRYPSVTSSQGSYYPGQASPGSQGQPGQEQPGQGGQGGQGGQ 153

QY 123 -----PPQGG-----GGGGGYYPCSRPGEVGYGQGGQ-----RQMYPPCRGTGGGP 168  
Db 154 YNPTSPQPGQGGQGGQGGQGGQGGQGGQGGQGGQGGQGGQGGQGGQGGQGGQ 213

QY 169 RIGR 172

Db 214 QPGQ 217

## RESULT 15

JC4966

high-molecular-weight glutenin - wheat

N:Alternate names: prolamine

C:Species: Triticum aestivum (common wheat)

C:Date: 31-Dec-1996 #sequence\_revision 31-Dec-1996 #text\_change 09-May-1997

C:Accession: JC4966

R:Anderson, O.D.; Kuhl, J.C.; Tam, A.

Gene 174, 51-58, 1996

A;Title: Construction and expression of a synthetic wheat storage protein gene.  
A;Reference number: JC4966; MUID:97017127; PMID:8863728  
A;Accession: JC4966  
A;Molecule type: DNA  
A;Residues: 1-161 <AND>  
A;Note: the authors translated the codon GAT for residue 12 and GAC for residue 31 as Glutamine.  
C;Comment: This protein belongs to one class of wheat prolamines.  
C;Superfamily: glutenin

Query Match 11.6%; Score 133.5; DB 2; Length 161;  
Best Local Similarity 27.3%; Pred. No. 0.00035;  
Matches 54; Conservative 19; Mismatches 64; Indels 61; Gaps 9;

QY 24 QGVEVERQLRDLCQ---WQEVQSPILDACRQVLDRLQLTGGGGGGVGGPPFRWGTGLRMRC 80  
Db 2 EGASEQ-----LQCDRELQELQERELKACQVMD----- 31

QY 81 QQLQDVSRRCRCAAIRSMVRGYEE--AMPPLKGMWPMGRQOQPPQ-----GGGGGGG 132  
Db 32 QQLRDISPECHPVVSVFVAGVQYEQIIVVP--KGGTFYPGETTPPQQLQRIFWGIPALL 89

QY 133 GYYPGCSRPGEGYGGGGGQGMYPGCRPTTGGGPRI-----GRVLTAKREYAAGLP 186  
Db 90 KRYTPSVTCPPQVSYTPG---QASPORSRDITSSSYHVSVEHQASLKVAKAQQLAALP 146

QY 187 MLCRLSEPPQECISIFSGD 204  
Db 147 MLCRL-----EGGD 155

Search completed: November 29, 2003, 17:07:05  
Job time : 29 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 29, 2003, 15:54:39 ; Search time 18 Seconds  
(without alignments)  
538.195 Million cell updates/sec

Title: US-10-053-410-4

Perfect score: 1148

Sequence: 1 MAKIAAAAAAALCFALVAV.....MMCRLEPQECISFGSDQY 206

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_41.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match %	Description
1	435	37.9	P29835 oryza sativ
2	289.5	25.2	P08488 triticum ae
3	281	24.5	P10387 triticum ae
4	156.5	13.6	P10388 triticum ae
5	151.5	13.2	P08489 triticum ae
6	114.5	10.0	P04403 bertholletii
7	114	9.9	P04724 triticum ae
8	112.5	9.8	P01089 ricinus com
9	112	9.8	P15461 helianthus
10	111.5	9.7	P18573 triticum ae
11	111	9.7	P04726 triticum ae
12	110	9.6	P02861 triticum ae
13	110	9.6	P16851 triticum ae
14	110	9.6	P48810 drosophila
15	109.5	9.5	P04722 triticum ae
16	109	9.5	Q01885 oryza sativ
17	106.5	9.3	P16850 triticum ae
18	106.5	9.3	P04727 triticum ae
19	105	9.1	P01088 zea mays lm
20	105	9.1	Q9H461 homo sapien
21	105	9.1	P33485 pseudorabie
22	103	9.0	Q61545 mus musculu
23	102.5	8.9	P23463 mus musculu
24	102	8.9	P01087 eleusine co
25	102	8.9	P01086 hordeum vul
26	102	8.9	P30845 triticum ae
27	101.5	8.8	Q63796 rattus norv
28	101	8.8	P10979 zea mays lm
29	101	8.8	P35085 dictyosteli
30	101	8.8	Q12906 h interleuk
31	100.5	8.8	P25074 oryza sativ
32	100.5	8.8	Q9bzw0 homo sapien
33	100.5	8.8	Q01860 homo sapien

Query Match 37.9%; Score 435; DB 1; Length 186;

34	100	8.7	304	1	GBD1_WHEAT	P04729 triticum ae
35	100	8.7	465	1	GRP2_PHAVU	P10496 phaseolus v
36	100	8.7	656	1	EW5_HUMAN	Q01844 homo sapien
37	99.5	8.7	145	1	IAA_HORVU	P28041 hordeum vul
38	99.5	8.7	212	1	EGG1_SCHJA	P19470 schistosoma
39	99.5	8.7	220	1	AVE3_AVESA	P80356 avena sativ
40	98.5	8.6	124	1	IAA1_WHEAT	P01085 triticum ae
41	98.5	8.6	289	1	HOG3_HORVU	P80198 hordeum vul
42	97.5	8.5	186	1	GOA8_WHEAT	P04728 triticum ae
43	97	8.4	169	1	GRP2_SINAL	P49311 sinapis alb
44	97	8.4	461	1	STR3_RAT	P58405 rattus norv
45	96.5	8.4	512	1	ANX7_XENLA	Q92125 xenopus lae

## ALIGNMENTS

RESULT 1  
GL19\_ORYSA  
ID GL19\_ORYSA STANDARD; PRT; 186 AA.  
AC P29835;  
DT 01-APR-1993 (Rel. 25, Created)  
DT 01-APR-1993 (Rel. 25, Last sequence update)  
DT 01-JUN-1994 (Rel. 29, Last annotation update)  
DE 19 kDa globulin precursor (Alpha-globulin).  
OS Oryza sativa (Rice).  
OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
OC Ehrhartoideae; Oryzoae; Oryza.  
OX NCBI\_TaxID=4530;  
RN [1]  
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.  
RC STRAIN=cv. Nipponbare; TISSUE=Endosperm;  
RX MEDLINE=92119226; PubMed=1731968;  
RA Shorrosh B.S., Wen L., Zen K.C., Huang J.K., Pan J.S., Hermodson M.A.,  
RA Tanaka K., Muthukrishnan S., Reeck G.R.;  
RT "A novel cereal storage protein: molecular genetics of the 19 kDa  
RT globulin of rice.";  
RL Plant Mol. Biol. 18:151-154(1992).  
RN [2]  
RP SEQUENCE OF 5-186 FROM N.A.  
RC STRAIN=cv. Lamont; TISSUE=Endosperm;  
RX MEDLINE=93277591; PubMed=8503935;  
RA Krishnan H.B., Pueppke S.G.;  
RT "Nucleotide sequence of an abundant rice seed globulin: homology with  
RT the high molecular weight glutelins of wheat, rye and triticale.";  
RL Biochem. Biophys. Res. Commun. 193:460-466(1993).  
CC -!- FUNCTION: SEED STORAGE PROTEIN.  
CC -!- SIMILARITY: BELONGS TO THE 2S SEED STORAGE ALBUMINS FAMILY.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; X63990; CAA45400.1; -  
CC EMBL; L12252; AAA72362.1; ALT\_INIT.  
CC PIR; S20024; WNR219.  
CC Gramene; P29835; -  
CC InterPro; IPR003612; AAI.  
CC InterPro; IPR001419; Glutenin.  
CC Pfam; PF00234; trypan\_alpha\_aml; 1.  
CC PRINTS; PR00210; GLUTENIN.  
CC SMART; SM00499; AAI; 1.  
CC Signal; Seed storage protein.  
CC SIGNAL 1 22 POTENTIAL.  
CC FT CHAIN 23 186 19 kDa GLOBULIN.  
CC SQ SEQUENCE 186 AA; 21050 MW; 9E09BA74C80B6810 CRC64;  
Query Match 37.9%; Score 435; DB 1; Length 186;

Best Local Similarity 51.0%; Pred. No. 7.7e-28;  
Matches 107; Conservative 20; Mismatches 55; Indels 28; Gaps 9;

QY 1 MAKIAAAAAALCFALVAVAVCQGEVERQRLDLQWQVQESPLDACRQVLDRLQLTGG 60  
DB 1 MAAVVAFAAL-MAAAMVAISGAHSESEMRFRDQCQREVQDSPLDACRQVLDRLQLTGR 59  
QY 61 GGGGGVGPFRWGTGLMRCCQQLQDYSRECRCAAIRSMVRGYEAMP-PLEKGMWPGRQ 119  
DB 60 ERPQPMFRFRPAGLGLMRCCQQLQDYSRECRCAAIRSMVRGYEAMP-PLEKGMWPGRQ 119  
QY 120 QPPPPGQGGGQGGYYPGSRPCEGYGGGQGGQOMY---PPCRPGTGTGGPRIGRVLRT 176  
DB 120 YY-----GGEG-----SSSQGY-YGEGSEEGYGEQQQPGMT-----RVRLT 158  
QY 177 KAREYAAGLPMCMRLSEPOCSIFSGDQY 206  
DB 159 RARQYAAQLPSMCRV-EPOQCSIFAAG-QY 186

## RESULT 2

GLT3\_WHEAT  
ID GLT3\_WHEAT STANDARD; PRT; 660 AA.  
AC P08488;  
DT 01-AUG-1988 (Rel. 08, Created)  
DT 01-AUG-1988 (Rel. 08, Last sequence update)  
DT 15-JUL-1999 (Rel. 38, Last annotation update)  
DE Glutenin, high molecular weight subunit 12 precursor.  
OS Triticum aestivum (Wheat).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae;  
OC Triticeae; Triticum.  
OX NCBI\_TaxID=4565;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=cv. Chinese Spring;  
RX MEDLINE=86041882; PubMed=3840588;  
RA Thompson R.D., Bartels D., Harberd N.P.;  
RT "Nucleotide sequence of a gene from chromosome 1D of wheat encoding a  
HMW-glutenin subunit";  
RL Nucleic Acids Res. 13:6833-6846(1985).  
CC -1- FUNCTION: GLUTENINS ARE THE HIGH MOLECULAR WEIGHT SEED STORAGE  
PROTEIN OF WHEAT ENDOSPERM. THOUGHT TO BE RESPONSIBLE FOR THE  
VISCO-ELASTIC PROPERTY OF WHEAT DOUGH.  
CC -1- SUBUNIT: DISULFIDE-BRIDGE LINKED AGGREGATES.  
CC -1- MISCELLANEOUS: GLUTENINS ARE CODED BY SEVERAL GENES ON EACH OF THE  
GROUP 1 CHROMOSOMES OF WHEAT.  
CC -1- MISCELLANEOUS: THE MATURE PROTEIN IS CHARACTERIZED BY A LARGE  
NUMBER OF WELL PRESERVED REPEATS OF THE TWO MOTIFS: GQPGQG AND  
GQPGQGQGGYPTS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC EMBL; X03041; CAA26847.1; -  
DR PIR; A24266; A24266.  
DR InterPro; IPR003612; AAI.  
DR Pfam; PF03157; Glutenin\_hmw; 1.  
DR PRINTS; PR00210; GLUTENIN.  
DR SMART; SM00499; AAI; 1.  
KW Seed storage protein; Repeat; Multigene family; Signal.  
FT SIGNAL 1 21 GLUTENIN, HIGH MOLECULAR WEIGHT SUBUNIT  
FT CHAIN 22 660

FT DOMAIN 125 615  
FT SEQUENCE 660 AA; 70868 MW; 2BFD09DB8C8FCC64; CRC64;

Query Match 25.2%; Score 289.5; DB 1; Length 660;  
Best Local Similarity 36.2%; Pred. No. 7.8e-16;  
Matches 80; Conservative 24; Mismatches 52; Indels 65; Gaps 10;

QY 1 MAKIAAAAAALCFALVAVAVCQGEVERQRLDLQWQVQESPLDACRQVLDRLQLTGG 60  
DB 1 MAKRLVLPAAVV--IALVALTTAEGEASRQ----LQERELQESSLEACRQVVDQQL--- 51  
QY 61 GGGGGVGPFRWGTGLMRCCQQLQDYSRECRCAAIRSMVRGYEAMP-PLEKGMWPGRQ 115  
DB 52 -----AGRLPMWGTGLMRCCQQLQDYSRECRCAAIRSMVRGYEAMP-PLEKGMWPGRQ 106  
QY 116 -----WG-----RQ-----QPPPPGQGGGQGG--- 133  
DB 107 PLQQLQQGIFWGTSSQTQGYYPSTVPSRQSGYYPGQASPPQPGQGGQPKWQEPGQGGQ 166  
QY 134 YYYPCS--RPGEYGYGGQGGQRYPPCRPGTGTGGPRIGR 172  
DB 167 WYPTSLQPGGQGGQIGKG--KQGYPTSLQPGGQGGQIGQ 205

## RESULT 3

GLT0\_WHEAT  
ID GLT0\_WHEAT STANDARD; PRT; 648 AA.  
AC P10387;  
DT 01-MAR-1989 (Rel. 10, Created)  
DT 01-MAR-1989 (Rel. 10, Last sequence update)  
DT 15-JUL-1999 (Rel. 38, Last annotation update)  
DE Glutenin, high molecular weight subunit DY10 precursor.  
GN GLU-D1-2B.  
OS Triticum aestivum (Wheat).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae;  
OC Triticeae; Triticum.  
OX NCBI\_TaxID=4565;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=cv. Cheyenne;  
RX MEDLINE=89098419; PubMed=2563152;  
RA Anderson O.D., Greene F.C., Yip R.E., Halford N.G., Shewry P.R.,  
RA Malpica-Romero J.M.;  
RT "Nucleotide sequences of the two high-molecular-weight glutenin genes  
from the D-genome of a hexaploid bread wheat, Triticum aestivum L. cv  
Cheyenne";  
RL Nucleic Acids Res. 17:461-462(1989).  
CC -1- FUNCTION: GLUTENINS ARE THE HIGH MOLECULAR WEIGHT SEED STORAGE  
PROTEIN OF WHEAT ENDOSPERM. THOUGHT TO BE RESPONSIBLE FOR THE  
VISCO-ELASTIC PROPERTY OF WHEAT DOUGH.  
CC -1- SUBUNIT: DISULFIDE-BRIDGE LINKED AGGREGATES.  
CC -1- MISCELLANEOUS: GLUTENINS ARE CODED BY SEVERAL GENES ON EACH OF THE  
GROUP 1 CHROMOSOMES OF WHEAT.  
CC -1- MISCELLANEOUS: THE MATURE PROTEIN IS CHARACTERIZED BY A LARGE  
NUMBER OF WELL PRESERVED REPEATS OF THE TWO MOTIFS: GQPGQG AND  
GQPGQGQGGYPTS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC EMBL; X12929; CAA31396.1; -  
DR PIR; S04832; S04832.  
DR InterPro; IPR003612; AAI.  
DR Pfam; PF03157; Glutenin\_hmw; 1.  
DR PRINTS; PR00210; GLUTENIN.  
DR SMART; SM00499; AAI; 1.  
KW Seed storage protein; Repeat; Multigene family; Signal.  
FT SIGNAL 1 21  
FT CHAIN 22 648 GLUTENIN, HIGH MOLECULAR WEIGHT SUBUNIT



```
DR Pfam; PF03157; Glutenin_hmw; 1.
DR PRINTS; PR00210; GLUTENIN
KW Seed storage protein; Repeat; Multigene family; Signal.
FT SIGNAL 1 21
FT CHAIN 22 838 GLUTENIN, HIGH MOLECULAR WEIGHT SUBUNIT
FT DOMAIN 130 799 REPEATS.
FT SEQUENCE 838 AA; 89174 MW; 71D715B7BDF0722D CRC64;
SQ
Query Match 13.2%; Score 151.5; DB 1; Length 838;
Best Local Similarity 28.7%; Pred. No. 7.6e-05;
Matches 58; Conservative 18; Mismatches 45; Indels 81; Gaps 10;
QY 1 MAKIAAAAAALCFALVAVAVCOGEVERQRLDLOC---WQVQSPDLDACQVLDRLQ 57
DB 1 MAKRLVLFVAV---VALVALTVAEGEASEQ---LQERELQELQRELKACQVMD--- 51
QY 58 TGGGGGGGGVFRWGTLGRMRCQQQLQDVSRCAIRSMVRGYEE----- 104
DB 52 -----QQLRDISPECHPVVSPVAGYEQIIVPKGGSFYPG 88
QY 105 -----AMPPLEKGNWP--WGRQQ-----QPPQGGGGGGGGYYPGSRPG 142
DB 89 ETTTPQLOQRIFWGIPALLKRYVPSVTSPQVSYYPGQASPORPGQGQ----- 140
QY 143 EGYGYGGGQRMYP--PCRPG 162
DB 141 QGQSGGQ-GQGGYPTSPQPG 161
RESULT 6
2SS_BEREX
ID_2SS_BEREX STANDARD; PRT; 146 AA.
AC P04403; P04402;
DT 20-MAR-1987 (Rel. 04, Created)
DT 01-OCT-1989 (Rel. 12, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE 2S sulfur-rich seed storage protein precursor (Allergen Ber e 1).
GN B2S1 AND B2S2.
OS Bertholletia excelsa (Brazil nut).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Asteridae; Ericales; Lecythidaceae; Bertholletia.
OX NCBI_TaxID=3645;
RN [1]
RP SEQUENCE FROM N.A.
RA Altenbach S.B., Pearson K.W., Leung F.W., Sun S.S.M.;
RT "Cloning and sequence analysis of a cDNA encoding a Brazil nut protein
exceptionally rich in methionine.";
RL Plant Mol. Biol. 8:239-250(1987).
RN [2]
RP SEQUENCE FROM N.A.
RA Bassuener R.;
RL Submitted (DEC-1990) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=91370890; PubMed=1840683;
RA Gander E.S., Holmstrom K.O., de Paiva G.R., de Castro L.A.B.,
RA Carneiro M., Grossi de Sa M.F.;
RT "Isolation, characterization and expression of a gene coding for a 2S
albumin from Bertholletia excelsa (Brazil nut).";
RL Plant Mol. Biol. 15:437-448(1991).
RN [4]
RP SEQUENCE OF 37-64 AND 70-142.
RX MEDLINE=87004679; PubMed=3759080;
RA Ampe C., van Damme J., de Castro L.A.B., Sampaio M.J.A.M.,
RA van Montagu M., Vandekerckhove J.;
RT "The amino-acid sequence of the 2S sulphur-rich proteins from seeds
of Brazil nut (Bertholletia excelsa H.B.K.).";
RL Eur. J. Biochem. 159:597-604(1986).
CC -|- FUNCTION: THIS IS A 2S SEED STORAGE PROTEIN.
CC -|- SUBUNIT: THE MATURE PROTEIN CONSISTS OF A SMALL AND A LARGE CHAIN
LINKED BY DISULFIDE BONDS.
```

"Evolution and heterogeneity of the alpha-/beta-type and gamma-type gliadin DNA sequences.";  
 J. Biol. Chem. 260:8203-8213(1985).  
 CC -(-) FUNCTION: GLIADIN IS THE MAJOR SEED STORAGE PROTEIN IN WHEAT.  
 CC -(-) MISCELLANEOUS: THE ALPHA/BETA-GLIADINS CAN BE DIVIDED INTO 5  
 CC HOMOLOGUE CLASSES. SEQUENCE DIVERGENCE BETWEEN THE CLASSES IS DUE  
 CC TO SINGLE BASE SUBSTITUTIONS AND TO DUPLICATIONS OR DELETIONS  
 CC WITHIN OR NEAR DIRECT REPEATS. THERE ARE MORE THAN A 100 COPIES OF  
 CC THE GENE FOR ALPHA/BETA-GLIADIN PER HAPLOID GENOME.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; M11075; AAA34282.1; -;  
 DR PIR; T06500; T06500.  
 DR InterPro; IPR003612; AAI.  
 DR Pfam; PF00234; tryp\_alpha\_aml; 1.  
 DR SMART; SM00208; GLIADGLUTEN.  
 DR SMART; SM00499; AAI; 1.  
 KW Seed storage protein; Repeat; Signal; Multigene family.  
 FT SIGNAL 1 20  
 FT CHAIN 21 297 ALPHA/BETA-GLIADIN A-IV  
 SQ SEQUENCE 297 AA; 34239 MW; 002SED89A89588B CRC64;  
 Query Match 9.9%; Score 114; DB 1; Length 297;  
 Best Local Similarity 23.0%; Pred. No. 0.026;  
 Matches 44; Conservative 29; Mismatches 76; Indels 42; Gaps 6;  
 QY 24 QGEVQRRLDLQWQVQSPDLACQVLDRLQTLGGGGGGVGFPMGTGLMRCCQQL 83  
 DB 125 QQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQ 184  
 QY 84 QDVSRRCRCARISMRVG-----YEEAMPPEKGMWPMGRQQQPPPGGGGG 130  
 DB 185 WQIPEQSRCAIHNHVHAIILHQQQQQQQQQQQQQQQQQQQQQQQQQQQQQ 238  
 QY 131 QGGVYPPSPGEGYGGYGGG-GRQMYPPCPGTTGGGPRIGRVLTAKAREYA-AGLPMW 188  
 DB 239 QGSF-----QPSQNPQAGSVQFQQLP-----QFEIENLALETIPAM 277  
 QY 189 CRLSEPOECI 199  
 DB 278 CNVYIPPYCTI 288  
 RESULT 8  
 ID 2SS\_RICCO STANDARD; PRT; 258 AA.  
 AC P01089;  
 DT 21-JUL-1986 (Rel. 01, Created)  
 DT 01-NOV-1990 (Rel. 16, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE 2S albumin precursor (Allergen Ric c 1).  
 OS Ricinus communis (Castor bean).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC eurosids 1; Malpighiales; Euphorbiaceae; Ricinus.  
 OX NCBI\_TaxID=3988;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Endosperm;  
 RX MEDLINE=91109729; PubMed=2274038;  
 RA Irwin S.D., Keen J.N., Findlay J.B.C., Lord J.M.;  
 RT "The Ricinus communis 2S albumin precursor: a single preproprotein  
 RL may be processed into two different heterodimeric storage proteins.";  
 RL Mol. Gen. Genet. 222:400-408(1990).  
 RN [2]

RP SEQUENCE FROM N.A.  
 RC TISSUE=Endosperm;  
 RX MEDLINE=91016940; PubMed=2216785;  
 RA Irwin S.D., Lord J.M.;  
 RT "Nucleotide sequence of a Ricinus communis 2S albumin precursor  
 RL gene.";  
 RL Nucleic Acids Res. 18:5890-5890(1990).  
 RN [3]  
 RP SEQUENCE OF 157-190 AND 194-258.  
 RX MEDLINE=83082772; PubMed=7174664;  
 RA Sharief F.S., Li S.S.-L.;  
 RT "Amino acid sequence of small and large subunits of seed storage  
 RL protein from Ricinus communis.";  
 RL J. Biol. Chem. 257:14753-14759(1982).  
 RN [4]  
 RP SIMILARITY TO PROTEINASE INHIBITORS.  
 RX MEDLINE=83308577; PubMed=6615448;  
 RA Odani S., Koide T., Ono T., Ohnishi K.;  
 RT "Structural relationship between barley (Hordeum vulgare) trypsin  
 RL inhibitor and castor-bean (Ricinus communis) storage protein.";  
 RL Biochem. J. 213:543-545(1983).  
 CC -(-) FUNCTION: THIS IS A 2S SEED STORAGE PROTEIN.  
 CC LINKED BY 2 DISULFIDE BONDS.  
 CC -(-) SUBUNIT: THE NATURE PROTEIN CONSISTS OF A SMALL AND A LARGE CHAIN  
 CC CHAINS INVOLVE CYS-162 AND CYS-175.  
 CC -(-) SIMILARITY: BELONGS TO THE 2S SEED STORAGE ALBUMINS FAMILY.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; X54158; CAA38097.1; -;  
 DR PIR; S11499; RZCS.  
 DR InterPro; IPR003612; AAI.  
 DR InterPro; IPR00617; Napin.  
 DR Pfam; PF00234; tryp\_alpha\_aml; 2.  
 DR PRINTS; PR00496; NAPIN.  
 DR SMART; SM00499; AAI; 1.  
 KW Seed storage protein; Signal; Allergen; Pyrrolidone carboxylic acid.  
 FT SIGNAL 1 21  
 FT PROPEP 22 156 2S ALBUMIN, SMALL CHAIN.  
 FT CHAIN 157 190  
 FT PROPEP 191 193 2S ALBUMIN, LARGE CHAIN.  
 FT CHAIN 194 258 PYRROLIDONE CARBOXYLIC ACID.  
 FT MOD.RES 194 194 E -> Q (IN REF. 3).  
 FT CONFLICT 222 222 MISSING (IN REF. 3).  
 FT CONFLICT 226 229 D -> N (IN REF. 3).  
 FT CONFLICT 234 234 E -> Q (IN REF. 3).  
 FT CONFLICT 255 255  
 SQ SEQUENCE 258 AA; 29290 MW; 27874CFC50E41072 CRC64;  
 Query Match 9.8%; Score 112.5; DB 1; Length 258;  
 Best Local Similarity 21.7%; Pred. No. 0.03;  
 Matches 59; Conservative 34; Mismatches 88; Indels 91; Gaps 11;  
 QY 1 MAKI-AAAAAALCF-----AALVAVACGVEVERLRDLQCMQVQESPLDA 48  
 DB 1 MAKLIPTIALVSVLLFIANASFAVTTTITTEIDSKGEREGSSSQCRQVQRDLSS 60  
 QY 49 CRQVLDRLQTLGGGGGGVGFPMGTGLRM-----RCCQQLQDVSRRCRC 94  
 DB 61 CERYLRQSSSRSSPGEV-----LRMPGDNQSQESQQLQQCCNCKVQVRDEQCEA 112  
 QY 95 IRLSWRGVEEAMPPEKGMWPMGRQQQPPPGGGGGGGYVPCSR-----PQEGYGGQ 149  
 DB 113 IKYI-----AEDQIQQQL-HGESEVAORAGEIVSSCGVRCMRQTRTNPSQCGRCG 165  
 QY 150 GGORQWYPPC---RGTGGGGP----- 169



```
Db 166 IQSQNLRCQCEYIKQVSGGRRSDNQBSLRGCCDHLKQMSQRCBGLRQALEQQQ 225
QY 170 -IQVR---ITKAREYAAGLPMCRJSEPOEC 197
Db 226 SQGLOQDVFEAFRTAANLPSMCGVS-PTEC 256

RESULT 9
2SS5 HELAN STANDARD; PRT; 295 AA.
AC P15461;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE 2S seed storage protein precursor (2S albumin storage protein).
GN HAG5.
OS Helianthus annuus (Common sunflower).
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Asteridae; campanulids; Asterales; Asteraceae; Asteroideae;
OC Heliantheae; Helianthus.
OX NCBI_TaxID=4232;
RN 1;
RP SEQUENCE FROM N.A., AND SEQUENCE OF 162-173.
RC STRAIN=cv. Giant grey stripe;
RX MEDLINE=88142538; PubMed=2830455;
RA Allen R.D., Cohen E.A., Vonder Haar R.A., Adams C.A., Ma D.P.,
RA Nessler C.L., Thomas T.L.;
RT "Sequence and expression of a gene encoding an albumin storage
  protein in sunflower.";
RL Mol. Gen. Genet. 210:211-218 (1987).
CC -!- FUNCTION: THIS IS A 2S SEED STORAGE PROTEIN.
CC -!- PTM: IT IS POSSIBLE THAT THE 38 KDA PRECURSOR IS CLEAVED INTO TWO
CC POLYPEPTIDES THAT ARE APPROXIMATELY THE SAME SIZE. THE MATURE
CC PROTEIN IS COMPOSED OF A SINGLE POLYPEPTIDE CONTAINING ONE OR MORE
CC INTRA-MOLECULAR DISULFIDE LINKAGES.
CC -!- SIMILARITY: BELONGS TO THE 2S SEED STORAGE ALBUMINS FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; X06410; CAA29699.1; -
CC DR PIR; S01062; S01062.
CC DR InterPro; IPR003612; AAI.
CC DR Pfam; PF00234; tryp_alpha_amyl; 2.
CC DR SMART; SM00499; AAI; 2.
CC DR Seed storage protein; Signal; Multigene family.
CC FT SIGNAL 1 20 POTENTIAL.
CC FT PROPEP 21 163 POTENTIAL.
CC FT CHAIN 162 295 2S SEED STORAGE PROTEIN.
CC SQ SEQUENCE 295 AA; 34071 MW; 8958A106805142A1 CRC64;

Query Match 9.8%; Score 112; DB 1; Length 295;
Best Local Similarity 22.0%; Pred. No. 0.037;
Matches 38; Conservative 21; Mismatches 54; Indels 60; Gaps 4;

QY 36 QCQWVEQESPLDACRQVLDRLTGCGGGGCVGFRWG-----TGLRMCCQQLQDVSR 89
Db 177 QCRETEIQRVGRCQRFVEQMQQSPRSTRPYQRRPGQQQQQQQQQQQQQQQQQQQQ 236
QY 90 CRCAAIRSVRGVEEAMPLEKGMWPGWGRQQPPQGGGGGQGYYPYPCSRPEGYGYG 149
Db 237 CHCEAIQEVARRVNRQ-----PQQQQQRRGQFGGQ----- 267
QY 150 GGQRMYPPCRPGTGGPRIGRVLTKAREYAAGLPMCRJSEPOECIFSG 202
Db 268 -----EMETARRVIQNLNPNQCDL-EVQQTCTTG 295

RESULT 10
GDA9 WHEAT STANDARD; PRT; 307 AA.
ID GDA9 WHEAT
AC P18573;
DT 01-NOV-1990 (Rel. 16, Created)
DT 01-NOV-1990 (Rel. 16, Last sequence update)
DT 01-NOV-1990 (Rel. 16, Last annotation update)
DE Alpha/beta-gliadin WMI precursor (prolamin).
OS Triticum aestivum (Wheat).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae;
OC Triticeae; Triticum.
OX NCBI_TaxID=4565;
RN 1;
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Chinese Spring; TISSUE=Endosperm;
RX MEDLINE=91346679; PubMed=2102865;
RA Garcia-Maroto F., Manana C., Garcia-Olmedo F., Carbonero P.;
RT "Nucleotide sequence of a cDNA encoding an alpha/beta-type gliadin
  from hexaploid wheat (Triticum aestivum).";
RL Plant Mol. Biol. 14:867-868 (1990).
CC -!- FUNCTION: GLIADIN IS THE MAJOR SEED STORAGE PROTEIN IN WHEAT.
CC -!- MISCELLANEOUS: THE ALPHA/BETA-GLIADINS CAN BE DIVIDED INTO 5
CC HOMOLOGY CLASSES. SEQUENCE DIVERGENCE BETWEEN THE CLASSES IS DUE
CC TO SINGLE BASE SUBSTITUTIONS AND TO DUPLICATIONS OR DELETIONS
CC WITHIN OR NEAR DIRECT REPEATS. THERE ARE MORE THAN A 100 COPIES OF
CC THE GENE FOR ALPHA/BETA-GLIADIN PER HAPLOID GENOME.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; X17361; CAA35238.1; -
CC DR PIR; S10015; S10015.
CC DR InterPro; IPR003612; AAI.
CC DR InterPro; IPR001954; GliA glutenin.
CC DR Pfam; PF00234; tryp_alpha_amyl; 1.
CC DR PRINTS; PR00208; GLIADGLUTEN.
CC DR SMART; SM00499; AAI; 1.
CC DR Seed storage protein; Repeat; Signal; Multigene family.
CC FT SIGNAL 1 20
CC FT CHAIN 21 307 ALPHA/BETA-GLIADIN WMI.
CC SQ SEQUENCE 307 AA; 35397 MW; 06C1858BD96F1E08 CRC64;

Query Match 9.7%; Score 111.5; DB 1; Length 307;
Best Local Similarity 23.4%; Pred. No. 0.043;
Matches 44; Conservative 29; Mismatches 76; Indels 39; Gaps 6;

QY 24 QCQEVERQLDLQCQWVEQESPLDACRQVLDRLTGCGGGGCVGFRWGTLRMCCQQL 83
Db 138 QQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQ 197
QY 84 QDVSRRCRAAIRSVRG-----YEEAMPLEKGMWPGWGRQQPPQGGGGGQGG 133
Db 198 WQIPQSRQCAIHNVHAILHQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQ 251
QY 134 YYPYPCSRPEGYGYGCGG--QRMYPPCRPGTGGGRVLTGRVLTAREYA--AGLPM 191
Db 252 F-----QFSQQNPQAGSVQPPQLP-----QFEIRNLALETLPAMCNV 290
QY 192 SEPQECST 199
Db 291 YIPPYCTI 298

RESULT 11
GDA6 WHEAT
```

ID GDA6\_WHEAT STANDARD; PRT; 296 AA.  
 AC F04726;  
 DT 13-AUG-1987 (Rel. 05, Created)  
 DT 13-AUG-1987 (Rel. 05, Last sequence update)  
 DT 15-JUL-1999 (Rel. 38, Last annotation update)  
 DE Alpha/beta-gliadin clone PW1215 precursor (Prolamin).  
 OS Triticum aestivum (Wheat).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae;  
 OC Triticeae; Triticum.  
 OX NCBI\_TaxID=4565;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=85242077; PubMed=38939304;  
 RA Summer-Smith M., Rafalski J.A., Sugiyama T., Stoll M., Soell D.;  
 RT "Conservation and variability of wheat alpha/beta-gliadin genes.";  
 RL Nucleic Acids Res. 13:3905-3916(1985).  
 CC -!- FUNCTION: GLIADIN IS THE MAJOR SEED STORAGE PROTEIN IN WHEAT.  
 CC -!- MISCELLANEOUS: THE ALPHA/BETA-GLIADINS CAN BE DIVIDED INTO 5  
 CC HOMOLOGUE CLASSES. SEQUENCE DIVERGENCE BETWEEN THE CLASSES IS DUE  
 CC TO SINGLE BASE SUBSTITUTIONS AND TO DUPLICATIONS OR DELETIONS  
 CC WITHIN OR NEAR DIRECT REPEATS. THERE ARE MORE THAN A 100 COPIES OF  
 CC THE GENE FOR ALPHA/BETA-GLIADIN PER HAPLOID GENOME.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
 CC or send an email to license@isb-sib.ch).  
 CC -----  
 DR EMBL; X02538; AAA26383.1; -.  
 DR EMBL; K03074; AAA34277.1; -.  
 DR PIR; S07361; S07361.  
 DR InterPro; IPR003612; AAI.  
 DR Pfam; PF00234; tryp\_alpha\_aml; 1.  
 DR PRINTS; PR00208; GLIADGLUTEN.  
 DR SMART; SM00499; AAI; 1.  
 KW Seed storage protein; Repeat; Signal; Multigene family.  
 FT SIGNAL 1 20  
 FT CHAIN 21 296 ALPHA/BETA-GLIADIN CLONE PW1215.  
 SQ SEQUENCE 296 AA; 33941 MW; A9BDF590AD40F135 CRC64;  
 -----  
 Query Match 9.7%; Score 111; DB 1; Length 296;  
 Best Local Similarity 23.2%; Pred. No. 0.045;  
 Matches 43; Conservative 26; Mismatches 76; Indels 40; Gaps 6;  
 QY 24 QGEVERQLRLDLCMQEVOESPLDACRQVLDRLQTLGGGGGGVGPFRWGTLRMRCQQQL 83  
 Db 130 QQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQ 189  
 QY 84 QDVSRRCRAAIRSMVRG---YEEAMPPLEKGMWPGWRRQQPPPPGGGGGGGGGGYPCSR 140  
 Db 190 WQIPESRCQAHHNVHAILHQQQ-----RQQPSSQ-----VSLQQ 227  
 QY 141 PGEIGYGGGQGGQWYPPRCPTGGTGG-----PRIGRVLTAKREYAAGLPMWRLSE 193  
 Db 228 PQQQVPSGGG-----FPQSQNPQAQGSVQPOLPQFEIRNALQT-----LPRMCNVI 279  
 QY 194 PQECS 198  
 Db 280 PPKCS 284  
 -----  
 RESULT 12  
 ID\_GLT1\_WHEAT STANDARD; PRT; 101 AA.  
 AC F02861;  
 DT 21-JUL-1986 (Rel. 01, Created)  
 DT 21-JUL-1986 (Rel. 01, Last sequence update)  
 DT 01-OCT-1994 (Rel. 30, Last annotation update)

DE Glutenin, high molecular weight subunit PC256 (Fragment).  
 OS Triticum aestivum (Wheat).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae;  
 OC Triticeae; Triticum.  
 OX NCBI\_TaxID=4565;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=162360386(1983).  
 RA Forde J., Forde B.G., Fry R.P., Kreis M., Shewry P.R., Miflin B.J.;  
 RT "Identification of barley and wheat cDNA clones related to the high-  
 RT M-r polypeptides of wheat gluten.";  
 RL FEBS Lett. 162:360-366(1983).  
 CC -!- FUNCTION: GLUTENINS ARE THE HIGH MOLECULAR WEIGHT SEED STORAGE  
 CC PROTEIN OF WHEAT ENDOSPERM. THOUGHT TO BE RESPONSIBLE FOR THE  
 CC VISCOS-ELASTIC PROPERTY OF WHEAT DOUGH.  
 CC -!- SUBUNIT: DISULFIDE-BRIDGE LINKED AGGREGATES  
 CC -!- MISCELLANEOUS: GLUTENINS ARE CODED BY SEVERAL GENES ON EACH OF THE  
 CC GROUP 1 CHROMOSOMES OF WHEAT.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
 CC or send an email to license@isb-sib.ch).  
 CC -----  
 DR EMBL; X00054; CAA24933.1; -.  
 DR PIR; A03352; BEWTL.  
 DR InterPro; IPR001419; Glutenin.  
 DR Pfam; PF03157; Glutenin\_hmw; 1.  
 DR PRINTS; PR00210; GLUTENIN.  
 KW Seed storage protein; Repeat; Multigene family.  
 FT NON\_TER 1  
 SQ SEQUENCE 101 AA; 10896 MW; EFD6ADDDPEED993 CRC64;  
 -----  
 Query Match 9.6%; Score 110; DB 1; Length 101;  
 Best Local Similarity 33.3%; Pred. No. 0.02;  
 Matches 36; Conservative 11; Mismatches 25; Indels 36; Gaps 6;  
 QY 117 GRQQQP-----PPQGGGGGGGGYVPCSS--RFGVGYGQGGGQORQWYPPRCPTGGP- 168  
 Db 4 GCGQGPQWQLPQQQQG-----YYPTSPQSGGQQLGQG--QQGYPTSPQSGGQQQ 56  
 QY 169 -----RIGRVLTAKREYAAGLPMWRLSEPCSECFISGSD 204  
 Db 57 GYDSPYHVSAAEHQAASLKVAKAQAQLPAMCRL-----EGGD 95  
 -----  
 RESULT 13  
 ID\_IA02\_WHEAT STANDARD; PRT; 145 AA.  
 AC P16851;  
 DT 01-AUG-1990 (Rel. 15, Created)  
 DT 01-MAR-1992 (Rel. 21, Last sequence update)  
 DT 01-MAR-1992 (Rel. 21, Last annotation update)  
 DE Alpha-amylase/trypsin inhibitor CM2 precursor (Chloroform/methanol-  
 DE soluble protein CM2).  
 OS Triticum aestivum (Wheat).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae;  
 OC Triticeae; Triticum.  
 OX NCBI\_TaxID=4565;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=cv. Agathe; TISSUE=Seed;  
 RX MEDLINE=91370875; PubMed=1893104;  
 RA Gautier M.F., Alary R., Lullien V., Joudrier P.;  
 RT "Nucleotide sequence of a cDNA clone encoding the wheat (Triticum  
 RT durum Desf.) CM2 protein.";  
 RL Plant Mol. Biol. 16:333-334(1991).  
 RN [2]  
 RP SEQUENCE OF 26-53.



Search completed: November 29, 2003, 17:05:29  
Job time : 23 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 29, 2003, 16:53:04 ; Search time 56 Seconds  
(without alignments)  
949.265 Million cell updates/sec

Title: US-10-053-410-4  
Perfect score: 1148  
Sequence: 1 MAKIAAAAAALCFALVAV.....MMCRILSEPOCSIFSGDQY 206

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database: SPTRMBL\_23.\*

- 1: sp\_archaea.\*
- 2: sp\_bacteria.\*
- 3: sp\_fungi.\*
- 4: sp\_human.\*
- 5: sp\_invertebrate.\*
- 6: sp\_mammal.\*
- 7: sp\_mhc.\*
- 8: sp\_organelle.\*
- 9: sp\_phage.\*
- 10: sp\_plant.\*
- 11: sp\_prodent.\*
- 12: sp\_virus.\*
- 13: sp\_vertebrate.\*
- 14: sp\_unclassified.\*
- 15: sp\_rvirus.\*
- 16: sp\_bacteriap.\*
- 17: sp\_archaeap.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	1148	100.0	206	10	Q946V3 zea mays (m
2	439	38.2	186	10	P93414 oryza sativ
3	414	36.1	226	10	Q8LKV8 aegilops ta
4	292.5	25.5	654	10	Q93XR5 aegilops cy
5	289	25.2	720	10	Q941J6 triticum ae
6	288.5	25.1	624	10	Q8LKV7 aegilops ta
7	287	25.0	705	10	Q03871 triticum ae
8	283	24.7	704	10	Q8LKV8 aegilops sp
9	282.5	24.6	475	10	Q40045 hordeum vul
10	282	24.6	648	10	Q38767 aegilops ta
11	281.5	24.5	707	10	Q40054 hordeum vul
12	277.5	24.2	196	10	Q8H0L6 triticum ae
13	276.5	24.1	655	10	Q8S3V7 aegilops um
14	275	24.0	713	10	Q941K7 secale cere
15	267.5	23.3	737	10	Q941L4 secale cere
16	267	23.3	666	10	Q93XR6 aegilops cy

17	267	23.3	707	10	Q941L6 secale cere
18	267	23.3	713	10	Q941J8 triticum ae
19	267	23.3	713	10	Q9SDM2 triticum ae
20	267	23.3	713	10	Q93WM1 secale cere
21	264.5	23.0	179	10	Q9M5N3 aegilops ve
22	260	22.6	633	10	Q8S3V9 aegilops ma
23	259	22.6	713	10	Q941K8 secale cere
24	257	22.4	181	10	Q8H0L3 triticum ae
25	257	22.4	587	10	Q93XJ7 triticum ti
26	253	22.0	713	10	Q941L1 secale cere
27	172.5	15.0	743	10	Q941L0 secale cere
28	166	14.5	754	10	Q93WFO secale cere
29	166	14.5	754	10	Q941K9 secale cere
30	165	14.4	754	10	Q941L2 secale cere
31	165	14.4	766	10	Q9SDM3 triticum ae
32	164	14.3	462	10	Q941J9 triticum ae
33	163.5	14.2	818	10	Q93XJ8 triticum ti
34	162	14.1	971	10	Q8GV12 triticum ae
35	160.5	14.0	238	10	Q8H0L2 triticum ae
36	160.5	14.0	811	10	Q941L5 secale cere
37	160.5	14.0	811	10	Q941J7 triticum ae
38	158.5	13.8	836	10	Q8S3T2 aegilops ta
39	158.5	13.8	836	10	Q8S3T3 aegilops ta
40	158	13.8	845	10	Q8LKV6 aegilops ta
41	157	13.7	204	10	Q9M5N2 aegilops ve
42	157	13.7	405	10	Q41516 triticum ae
43	157	13.7	405	10	Q9SYV0 triticum ae
44	157	13.7	766	10	Q941L3 secale cere
45	157	13.7	830	10	Q03872 triticum ae

#### ALIGNMENTS

#### RESULT 1

Q946V3 PRELIMINARY; PRT; 206 AA.  
ID Q946V3;  
AC Q946V3;  
DT 01-DEC-2001 (TREMELrel. 19, Created)  
DT 01-DEC-2001 (TREMELrel. 19, Last sequence update)  
DT 01-MAR-2003 (TREMELrel. 23, Last annotation update)  
DE Alpha globulin.  
OS Zea mays (Maize).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
OC PACCNAD clade; Panicoideae; Andropogoneae; Zea.  
OX NCBI\_TaxID=4577;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Endosperm;  
RX MEDLINE=21480069; PubMed=11595803;  
RA Woo Y.M., Hu D.W., Larkins B.A., Jung R.;  
RT "Genomics analysis of genes expressed in maize endosperm identifies novel seed proteins and clarifies patterns of zein gene expression.";  
RL Plant Cell 13:2297-2317(2001).  
DR EMBL; AF371278; RAL16993.1;  
DR InterPro; IPR003612; AAI.  
DR Fram; PF00234; tryp\_alpha\_amyl; 1.  
DR SMART; SM00499; AAI; 1.  
SQ SEQUENCE 206 AA; 22298 MW; B146CCA41BD82089 CRC64;

Query Match 100.0%; Score 1148; DB 10; Length 206;  
Best Local Similarity 100.0%; Pred. No. 1.5e-105;  
Matches 206; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MAKIAAAAAALCFALVAVAVCQGEVERQRDLQCWQEVESPLDQACRQLDRLTGG 60

DB 1 MAKIAAAAAALCFALVAVAVCQGEVERQRDLQCWQEVESPLDQACRQLDRLTGG 60

QY 61 GGGGGVGPFRWGTGLRMRCRCCQQLQDVSRRCRAAIRSMVRGYEAMPPLEKGMWPGRRQ 120

DB 61 GGGGGVGPFRWGTGLRMRCRCCQQLQDVSRRCRAAIRSMVRGYEAMPPLEKGMWPGRRQ 120







RN	[2]
RP	SEQUENCE FROM N.A.
RC	STRAIN=cv, Cheyenne; TISSUE=Endosperm;
RA	Halford N.G., Forde J., Anderson O.D., Greene P.C., Shewry P.R.;
RT	"The nucleotide and deduced amino acids sequences of an HMW glutenin
RT	subunit gene from chromosome 1B of bread wheat ( <i>Triticum aestivum</i> L.)
RT	and comparison with those of genes from chromosomes 1A and 1D."
RL	Theor. Appl. Genet. 75:117-126(1987).
DR	EMBL; X61026; CAA43361.1; -
DR	InterPro; IPR0031612; AAI.
DR	InterPro; IPR001419; Glutenin.
DR	Pfam; PF03157; Glutenin_hmw; 1.
DR	PRINTS; PR00210; GLUTENIN.
DR	SMART; SM00499; AAI; 1.
KW	Signal.
FT	SIGNAL
FT	CHAIN
FT	SEQUENCE
FT	1 21 POTENTIAL.
FT	22 705 HMW GLUTENIN SUBUNIT 1BY9.
FT	SEQUENCE 705 AA; 75702 MW; EBBC74594071OFF8 CRC64;
FT	Query Match 25.0%; Score 287; DB 10; Length 705;
FT	Best Local Similarity 32.9%; Pred. No. 6.8e-20;
FT	Matches 85; Conservative 24; Mismatches 47; Indels 102; Gaps 13;
QY	17 LVAVACQGVEVERQRURDLQCWQVQSPLDACRQVLDRQLTGGGCGGVGPFRWGTLGR-76     :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   15 LVALTAAEGEASRQ----LQCRERLOESSLEACRQVVDQQL-----AGRLPWSTGLQ 62
DB	
QY	77 MRCCQQLQDVSRRCRCAAIRSMVRGYEE-AMPLEKGMWP-----WGR--- 118       :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   63 MRCCQQLRVSAKRPRVAVSQVRQYEQTVPFKGGSFYPGETTPLOOLQQVIFWGTTSQ 122
DB	
QY	119 -----QQQP-----PPQGG-----GGGGGGY----- 134 123 TVQGYPSVSSPQGGPYYPQASPPQGGQPGKWQELGGQGYPTSLHSGSGCQGG 182
DB	
QY	135 YYPSC--RPGEYGYGG-----GORQMYP--PCRPG---TTGGG 167     :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   183 YPSSLQQPQGGQIGQCGQGYPTSLQPGQGIQGGQGYPTSPQHGPQRQPGQ 242
DB	
QY	168 PRIGR-VRLTKAREYAAG 184
DB	
QY	243 QQIGGQQLGQGRQIGQG 260
DB	
RESULT 8	
Q8LK18	
ID	Q8LK18 PRELIMINARY; PRT; 704 AA.
AC	Q8LK18;
DT	01-OCT-2002 (TrEMBLrel. 22, Created)
DT	01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT	01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE	HMW glutenin subunit y.
GN	GLU-LISY.
OS	Aegilops speltoides (Goat grass).
OC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC	Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae;
OC	Triticeae; Aegilops.
OX	NCBI_TaxID=4573;
RN	[1]
RN	SEQUENCE FROM N.A.
RA	De Bustos A., Corredor M., Jouve N.S.;
RP	"Characterization of HMW glutensins in diploid species of the genus
RT	Aegilops.";
RL	Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.
DR	EMBL; AF513640; AAM47280.1; -
DR	InterPro; IPR001419; Glutenin.
DR	Pfam; PF03157; Glutenin_hmw; 4.
DR	PRINTS; PR00210; GLUTENIN.
QY	SEQUENCE 704 AA; 75932 MW; 1A4B73EC03C3F32C CRC64;
QY	Query Match 24.7%; Score 283; DB 10; Length 704;
QY	Best Local Similarity 34.8%; Pred. No. 1.7e-19;
QY	Matches 77; Conservative 21; Mismatches 45; Indels 78; Gaps 9;

```

ID Q38767 PRELIMINARY; PRT; 648 AA.
AC Q38767;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE high molecular weight glutenin subunit.
GN GLU-1-2.
OS Aegilops tauschii (Aegilops squarrosa).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae;
OC Triticeae; Aegilops.
OX NCBI_TaxID=37682;
RN [1]
RN SEQUENCE FROM N.A.
RA Mackie A.M., Sharp P.J., Lagudah E.S.;
RT "The nucleotide and derived amino acid sequence of a HMW glutenin gene
RT from Triticum tauschii and comparison with those from the D genome of
RT bread wheat.";
RL J. Cereal Sci. 0:0-0(1996).
RN [2]
RN SEQUENCE FROM N.A.
RA Mackie A.M., Lagudah E.S., Sharp P.J., Lafandra D.;
RT "Molecular and biochemical characterisation of HMW glutenin subunits
RT from T. tauschii and the D genome of hexaploid wheat.";
RL J. Cereal Sci. 0:0-0(1996).
DR EMBL; U39229; AAC15244.1; -.
DR InterPro; IPR001612; AAI.
DR Pfam; PF03157; Glutenin_hmw; 1.
DR PRINTS; PR00210; GLUTENIN.
DR SMART; SM00499; AAI; 1.
SQ SEQUENCE 648 AA; 69567 MW; E5D71332E520DF5E CRC64;

Query Match 24.5%; Score 282; DB 10; Length 648;
Best Local Similarity 37.1%; Pred. No. 1.9e-19;
Matches 79; Conservative 22; Mismatches 46; Indels 66; Gaps 11;

QY 1 MAKIAAAAAALCFALVAVAVCGEVRQLDLCWQVQVPSPLDACKROVLDRLTGG 60
DB 1 MAKRLVLFVAVV--IALVALTTAEASRQ---LQERLQSSLEACRQVVDQL--- 51

QY 61 GGCGGGVGFPRFWGTGLRMRCQQQLQDVSRRCRAAIRSMVRGYEE--AMPPLEKGMWP--- 115
DB 52 -----AGRLPWSGTGLQMRCCQQLRDVSAKRSVAISQVARYEQTVVPPKGGSYFPGETT 106
QY 116 -----WG-----RQ-----QQPPPGGGGGGGG----- 133
DB 107 PLQLOOQGFIFWGTSSOTVGYGYPGVTSRQGSYPGQASPPQPGQGGQPKWQEPGQGGQ 166
QY 134 YYPCPS--RPEGVGVGQGGGQRMYPCC--RPG 162
DB 167 WYPTSLQPGGQQQIGK-GQGGYFTSLQPG 198

RESULT 11
Q40054 PRELIMINARY; PRT; 707 AA.
AC Q40054;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE hordein precursor.
GN HOR3.
OS Hordeum vulgare (Barley).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae;
OC Triticeae; Hordeum.
OX NCBI_TaxID=4513;
RN [1]
RN SEQUENCE FROM N.A.
RA Hirota N., Kuroda H., Ito K.;
RT Submitted (JAN-1996) to the EMBL/GenBank/DBJ databases.

Query Match 24.2%; Score 277.5; DB 10; Length 196;
Best Local Similarity 36.7%; Pred. No. 1.4e-19;
Matches 76; Conservative 22; Mismatches 44; Indels 65; Gaps 10;

QY 1 MAKIAAAAAALCFALVAVAVCGEVRQLDLCWQVQVPSPLDACKROVLDRLTGG 60
DB 1 MAKRLVLFVAVV--IALVALTTAEASRQ---LQERLQSSLEACRQVVDQL--- 51

QY 61 GGCGGGVGFPRFWGTGLRMRCQQQLQDVSRRCRAAIRSMVRGYEE--AMPPLEKGMWP--- 115
DB 52 -----AGRLPWSGTGLQMRCCQQLRDVSAKRSVAVSQVARYEQTVVPPKGGSYFPGETT 106
QY 116 -----WG-----RQ-----QQPPPGGGGGGGG----- 133
DB 107 PLQLOOQGFIFWGTSSOTVGYGYPGVTSRQGSYPGQASPPQPGQGGQPKWQEPGQGGQ 166
QY 134 YYPCPS--RPEGVGVGQGGGQRMYPCC--RPG 162
DB 167 WYPTSLQPGGQQQIGK-GQGGYFTSLQPG 198

RESULT 12
Q8H0L6 PRELIMINARY; PRT; 196 AA.
ID Q8H0L6;
AC Q8H0L6;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Glutenin high molecular weight subunit (Fragment).
GN HMW-GS (DY TYPE).
OS Triticum aestivum (Wheat).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae;
OC Triticeae; Triticum.
OX NCBI_TaxID=4565;
RN [1]
RN SEQUENCE FROM N.A.
RA Fan S., Guo A., Luo D., Zhao H., Feng X.;
RT "Partial sequence of wheat HMW-GS (Dy type) gene from cultivar 'Xiaoyan
RT 6'.";
RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ308967; CAC84118.1; -.
FT NON TER 196
SQ SEQUENCE 196 AA; 21587 MW; 4B8BE38F40A4D07 CRC64;

Query Match 24.2%; Score 277.5; DB 10; Length 196;
Best Local Similarity 36.7%; Pred. No. 1.4e-19;
Matches 76; Conservative 22; Mismatches 44; Indels 65; Gaps 10;

QY 1 MAKIAAAAAALCFALVAVAVCGEVRQLDLCWQVQVPSPLDACKROVLDRLTGG 60
DB 1 MAKRLVLFVAVV--IALVALTTAEASRQ---LQERLQSSLEACRQVVDQL--- 51

QY 61 GGCGGGVGFPRFWGTGLRMRCQQQLQDVSRRCRAAIRSMVRGYEE--AMPPLEKGMWP--- 115
DB 52 -----AGRLPWSGTGLQMRCCQQLRDVSAKRSVAVSQVARYEQTVVPPKGGSYFPGETT 106
QY 116 -----WG-----RQ-----QQPPPGGGGGGGG----- 133
DB 107 PLQLOOQGFIFWGTSSOTVGYGYPGVTSRQGSYPGQASPPQPGQGGQPKWQEPGQGGQ 166
```

```

DR EMBL; D82941; BAAL1642.1; -.
DR InterPro; IPR003612; AAI.
DR Pfam; PF03157; Glutenin_hmw; 1.
DR PRINTS; PR00210; GLUTENIN.
DR SMART; SM00499; AAI; 1.
KW Signal.
FT SIGNAL 1 28 POTENTIAL.
FT CHAIN 29 707 D HORDEIN.
SQ SEQUENCE 707 AA; 75108 MW; EBB873AF035D5789 CRC64;

Query Match 24.5%; Score 281.5; DB 10; Length 707;
Best Local Similarity 36.1%; Pred. No. 2.4e-19;
Matches 79; Conservative 22; Mismatches 51; Indels 67; Gaps 11;

QY 1 MAKIAAAAAALCFALVAVAVCGEVE-----RQLRLDLCWQVQVPSPLDACKROVLDR 54
DB 1 MAKRLVLFVAVV--IALVALTTAEAREINNNIFLDSRQLQERLQSSLEACRRVVD 58
QY 55 RQLTGGGGGGVGFPRFWGTGLRMRCQQQLQDVSRRCRAAIRSMVRGYEE----- 104
DB 59 QQL-----VGQLPWSGTGLQMRCCQQLRDVSPCRPVALSQVVRQYEQTEVPKSGGS 110
QY 105 -----AMPPLEK-GMW-----PMGRQ-----QPPPGGGGGGGG 133
DB 111 FYPGGTAPPLQGGGWMGTSVKWYYPDQTSQQSQWQGGYHQSIVTSQQPGQGGQSGYPG 170
QY 134 YYPCPSRPEGVGYGQGGGQRMYP-----PCRPGTTGG 166
DB 171 STFP-QQPGGGQ---QPGQRPMSYPSATFPQPGGGGG 205

RESULT 12
Q8H0L6 PRELIMINARY; PRT; 196 AA.
ID Q8H0L6;
AC Q8H0L6;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Glutenin high molecular weight subunit (Fragment).
GN HMW-GS (DY TYPE).
OS Triticum aestivum (Wheat).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae;
OC Triticeae; Triticum.
OX NCBI_TaxID=4565;
RN [1]
RN SEQUENCE FROM N.A.
RA Fan S., Guo A., Luo D., Zhao H., Feng X.;
RT "Partial sequence of wheat HMW-GS (Dy type) gene from cultivar 'Xiaoyan
RT 6'.";
RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ308967; CAC84118.1; -.
FT NON TER 196
SQ SEQUENCE 196 AA; 21587 MW; 4B8BE38F40A4D07 CRC64;

Query Match 24.2%; Score 277.5; DB 10; Length 196;
Best Local Similarity 36.7%; Pred. No. 1.4e-19;
Matches 76; Conservative 22; Mismatches 44; Indels 65; Gaps 10;

QY 1 MAKIAAAAAALCFALVAVAVCGEVRQLDLCWQVQVPSPLDACKROVLDRLTGG 60
DB 1 MAKRLVLFVAVV--IALVALTTAEASRQ---LQERLQSSLEACRQVVDQL--- 51

QY 61 GGCGGGVGFPRFWGTGLRMRCQQQLQDVSRRCRAAIRSMVRGYEE--AMPPLEKGMWP--- 115
DB 52 -----AGRLPWSGTGLQMRCCQQLRDVSAKRSVAVSQVARYEQTVVPPKGGSYFPGETT 106
QY 116 -----WG-----RQ-----QQPPPGGGGGGGG----- 133
DB 107 PLQLOOQGFIFWGTSSOTVGYGYPGVTSRQGSYPGQASPPQPGQGGQPKWQEPGQGGQ 166
```

```
QY 134 YYPCS--RPEGVGYGGGQORQWYPP 158
Db 167 WYPTSLQPGGQQIGKG--KQGYYP 191

RESULT 13
Q883V7 PRELIMINARY; PRT; 655 AA.
AC Q883V7;
DT 01-JUN-2002 (TReMBLrel. 21, Created)
DT 01-JUN-2002 (TReMBLrel. 21, Last sequence update)
DT 01-OCT-2002 (TReMBLrel. 22, Last annotation update)
DE HMW glutenin subunit.
GN GLU-1UY.
OS Aegilops umbellulata.
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae;
OC Triticeae; Aegilops.
OX NCBI_TaxID=4491;
RN [1]
RP SEQUENCE FROM N.A.
RA Liu Z., Yan Z., Wan Y., Liu K., Zheng Y., Wang D.;
RT "Analysis of HMW glutenin subunits and their coding sequences in two
RT diploid Aegilops species."
RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL: AP476962; AAL82616.1; -.
DR InterPro: IPR001419; Glutenin.
DR Pfam: PF03157; Glutenin_hmw; 1.
SQ SEQUENCE 655 AA; 70458 MW; 20FOA70E4D54FF2D CRC64;

Query Match 24.1%; Score 276.5; DB 10; Length 655;
Best Local Similarity 36.3%; Pred. No. 6.8e-19;
Matches 77; Conservative 21; Mismatches 45; Indels 69; Gaps 10;

QY 16 ALVAVAVCGEVEERQRLDLQCWQVESPLDRCQVLDRLQLTGGGGGVPFRWGTL 75
Db 14 ALVALTTAGEASRQ-----LQCERELQESLEACRQVVDQQL-----AGRLPWSTGL 61

QY 76 RMRCCQQLQDVSRCCRAIRSMVRGYEE-AMPPLEKGMWP-----WGROQ 120
Db 62 QMRCCQQLRDVSAKCRHVAVSQVARYEQTVPPKGSFYPGTTPLOQLQQGIFWGTSS 121

QY 118 -----RQ-----QQPPGGGGGGY-----YPPCS--RPEGY 146
Db 122 QTVOGYPSVTSRQSYYPGQASPPQGGQPGQKQWBPQPGQGYPTSLQOPVGGQ 181

QY 147 YGGGQRMYPPCRPGTTG-----GGPRIGR 172
Db 182 IGQG--QGGYPTSPQHTGQRQPVGGQIGQ 211

RESULT 14
Q94IK7 PRELIMINARY; PRT; 713 AA.
AC Q94IK7;
DT 01-DEC-2001 (TReMBLrel. 19, Created)
DT 01-DEC-2001 (TReMBLrel. 19, Last sequence update)
DT 01-MAR-2003 (TReMBLrel. 23, Last annotation update)
DE High molecular weight glutenin subunit y precursor.
GN GLU-1R.
OS Secale cereale (Rye).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae;
OC Triticeae; Secale.
OX NCBI_TaxID=4550;
RN [1]
RP SEQUENCE FROM N.A.
RA de Bustos A., Jouve N.;
RT "Characterization and analysis of new subunits-like wheat HMW-glutenin
RT encoded by Glu-1R gene of Secale cereale."
RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
RN [2]
```

```
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Imperial;
RA de Bustos A., Rubio P., Jouve N.;
RT "Characterization of two gene subunits on the 1R chromosome of rye as
RT orthologs of each of the Glu-1 genes of hexaploid wheat."
RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AJ314781; CAC40682.1; -.
DR InterPro: IPR001419; Glutenin.
DR Pfam: PF03157; Glutenin_hmw; 1.
DR PRINTS; PR00210; GLUTENIN.
KW SIGNAL.
FT CHAIN 1 21 HMW GLUTENIN SUBUNIT Y.
ET CHAIN 22 713
SQ SEQUENCE 713 AA; 76274 MW; 49EEEB182BBC19C0 CRC64;

Query Match 24.0%; Score 275; DB 10; Length 713;
Best Local Similarity 32.1%; Pred. No. 1.1e-18;
Matches 78; Conservative 21; Mismatches 46; Indels 98; Gaps 10;

QY 16 ALVAVAVCGEVEERQRLDLQCWQVESPLDRCQVLDRLQLTGGGGGVPFRWGTL 75
Db 14 ALVALTTAGEASRQ-----LQCERELQESLEACRQVVDQQL-----AGRLPWSTGL 61

QY 76 RMRCCQQLQDVSRCCRAIRSMVRGYEE-AMPPLEKGMWP-----WGROQ 120
Db 62 QMRCCQQLRDVSAKCRHVAVSQVARYEQTVPPKGSFYPGTTPLOQLQQGIFWGTSS 121

QY 121 Q-----PPQGG-----GGGGGY----- 135
Db 122 QTVOGYPSVTSRQSYYPGQASPPQGGQPGQKQWBPQPGQGYPTSLQOPVGGQ 181

QY 136 --YPCS--RPEGY-----GYGGGQ-----RQWYPCRPTTGGGR 169
Db 182 GHYPASQQPGGQGGHYPPASLQPPGQGGQGHYPASLQPPGQGGQTEPQGMQPGGQ 241

QY 170 IGR 172
Db 242 IQQ 244

RESULT 15
Q94IL4 PRELIMINARY; PRT; 737 AA.
AC Q94IL4;
DT 01-DEC-2001 (TReMBLrel. 19, Created)
DT 01-DEC-2001 (TReMBLrel. 19, Last sequence update)
DT 01-MAR-2003 (TReMBLrel. 23, Last annotation update)
DE High molecular weight glutenin subunit y precursor.
GN GLU-1R.
OS Secale cereale (Rye).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae;
OC Triticeae; Secale.
OX NCBI_TaxID=4550;
RN [1]
RP SEQUENCE FROM N.A.
RA de Bustos A., Jouve N.;
RT "Characterization and analysis of new subunits-like wheat HMW-glutenin
RT encoded by Glu-1R gene of Secale cereale."
RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
RN [2]
```

```
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Imperial;
RA de Bustos A., Rubio P., Jouve N.;
RT "Characterization of two gene subunits on the 1R chromosome of rye as
RT orthologs of each of the Glu-1 genes of hexaploid wheat."
RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AJ314781; CAC40682.1; -.
DR InterPro: IPR001419; Glutenin.
DR Pfam: PF03157; Glutenin_hmw; 1.
DR PRINTS; PR00210; GLUTENIN.
KW SIGNAL.
FT CHAIN 1 21 HMW GLUTENIN SUBUNIT Y.
ET CHAIN 22 713
SQ SEQUENCE 713 AA; 76274 MW; 49EEEB182BBC19C0 CRC64;

Query Match 24.0%; Score 275; DB 10; Length 713;
Best Local Similarity 32.1%; Pred. No. 1.1e-18;
Matches 78; Conservative 21; Mismatches 46; Indels 98; Gaps 10;

QY 16 ALVAVAVCGEVEERQRLDLQCWQVESPLDRCQVLDRLQLTGGGGGVPFRWGTL 75
Db 14 ALVALTTAGEASRQ-----LQCERELQESLEACRQVVDQQL-----AGRLPWSTGL 61

QY 76 RMRCCQQLQDVSRCCRAIRSMVRGYEE-AMPPLEKGMWP-----WGROQ 120
Db 62 QMRCCQQLRDVSAKCRHVAVSQVARYEQTVPPKGSFYPGTTPLOQLQQGIFWGTSS 121

QY 121 Q-----PPQGG-----GGGGGY----- 135
Db 122 QTVOGYPSVTSRQSYYPGQASPPQGGQPGQKQWBPQPGQGYPTSLQOPVGGQ 181

QY 136 --YPCS--RPEGY-----GYGGGQ-----RQWYPCRPTTGGGR 169
Db 182 GHYPASQQPGGQGGHYPPASLQPPGQGGQGHYPASLQPPGQGGQTEPQGMQPGGQ 241

QY 170 IGR 172
Db 242 IQQ 244

RESULT 15
Q94IL4 PRELIMINARY; PRT; 737 AA.
AC Q94IL4;
DT 01-DEC-2001 (TReMBLrel. 19, Created)
DT 01-DEC-2001 (TReMBLrel. 19, Last sequence update)
DT 01-MAR-2003 (TReMBLrel. 23, Last annotation update)
DE High molecular weight glutenin subunit y precursor.
GN GLU-1R.
OS Secale cereale (Rye).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae;
OC Triticeae; Secale.
OX NCBI_TaxID=4550;
RN [1]
RP SEQUENCE FROM N.A.
RA de Bustos A., Rubio P., Jouve N.;
RT "Characterization and analysis of new subunits-like wheat HMW-glutenin
RT encoded by Glu-1R gene of Secale cereale."
RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
RN [2]
```

```
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Imperial;
RA de Bustos A., Rubio P., Jouve N.;
RT "Characterization of two gene subunits on the 1R chromosome of rye as
RT orthologs of each of the Glu-1 genes of hexaploid wheat."
RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AJ314781; CAC40682.1; -.
DR InterPro: IPR001419; Glutenin.
DR Pfam: PF03157; Glutenin_hmw; 1.
DR PRINTS; PR00210; GLUTENIN.
KW SIGNAL.
FT CHAIN 1 21 HMW GLUTENIN SUBUNIT Y.
ET CHAIN 22 737
SQ SEQUENCE 713 AA; 76274 MW; 49EEEB182BBC19C0 CRC64;

Query Match 24.0%; Score 275; DB 10; Length 713;
Best Local Similarity 32.1%; Pred. No. 1.1e-18;
Matches 78; Conservative 21; Mismatches 46; Indels 98; Gaps 10;

QY 16 ALVAVAVCGEVEERQRLDLQCWQVESPLDRCQVLDRLQLTGGGGGVPFRWGTL 75
Db 14 ALVALTTAGEASRQ-----LQCERELQESLEACRQVVDQQL-----AGRLPWSTGL 61

QY 76 RMRCCQQLQDVSRCCRAIRSMVRGYEE-AMPPLEKGMWP-----WGROQ 120
Db 62 QMRCCQQLRDVSAKCRHVAVSQVARYEQTVPPKGSFYPGTTPLOQLQQGIFWGTSS 121

QY 121 Q-----PPQGG-----GGGGGY----- 135
Db 122 QTVOGYPSVTSRQSYYPGQASPPQGGQPGQKQWBPQPGQGYPTSLQOPVGGQ 181

QY 136 --YPCS--RPEGY-----GYGGGQ-----RQWYPCRPTTGGGR 169
Db 182 GHYPASQQPGGQGGHYPPASLQPPGQGGQGHYPASLQPPGQGGQTEPQGMQPGGQ 241

QY 170 IGR 172
Db 242 IQQ 244

RESULT 15
Q94IL4 PRELIMINARY; PRT; 737 AA.
AC Q94IL4;
DT 01-DEC-2001 (TReMBLrel. 19, Created)
DT 01-DEC-2001 (TReMBLrel. 19, Last sequence update)
DT 01-MAR-2003 (TReMBLrel. 23, Last annotation update)
DE High molecular weight glutenin subunit y precursor.
GN GLU-1R.
OS Secale cereale (Rye).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae;
OC Triticeae; Secale.
OX NCBI_TaxID=4550;
RN [1]
RP SEQUENCE FROM N.A.
RA de Bustos A., Rubio P., Jouve N.;
RT "Characterization and analysis of new subunits-like wheat HMW-glutenin
RT encoded by Glu-1R gene of Secale cereale."
RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
RN [2]
```



**THIS PAGE BLANK (USPTO)**

Result No.	Query			DB	ID	Description
	Score	Match	Length			
1	130.5	11.4	157	3	US-08-706-391B-7	Sequence 7, Appli
2	121.5	10.0	308	4	US-09-252-991A-27249	Sequence 27249, A
3	114.5	10.0	410	4	US-09-252-991A-25812	Sequence 25812, A
4	109.5	9.5	426	1	US-07-918-023-2	Sequence 2, Appli
5	103	9.0	476	4	US-09-252-991A-17887	Sequence 17887, A
6	101.5	8.8	863	4	US-09-252-991A-19574	Sequence 19574, A
7	100.5	8.8	535	4	US-09-252-991A-28410	Sequence 28410, A
8	100	8.7	209	4	US-09-252-991A-30648	Sequence 30648, A
9	100	8.7	656	2	US-08-343-443B-2	Sequence 4, Appli
10	100	8.7	656	3	US-09-214-564A-4	Sequence 2, Appli
11	99	8.6	238	4	US-08-216-846-2	Sequence 17915, A
12	98.5	8.6	124	1	US-08-216-846-2	Sequence 2, Appli
13	98.5	8.6	124	1	US-08-782-177-3	Sequence 3, Appli
14	98.5	8.6	569	4	US-09-252-991A-24513	Sequence 24513, A
15	98.5	8.6	639	4	US-09-252-991A-28453	Sequence 28453, A
16	97.5	8.5	393	4	US-09-252-991A-30202	Sequence 30202, A
17	97	8.4	148	1	US-08-207-904-15	Sequence 15, Appl
18	96.5	8.4	188	4	US-09-252-991A-24358	Sequence 24358, A
19	96.5	8.4	493	3	US-08-556-978B-59	Sequence 59, Appl
20	96.5	8.4	494	4	US-09-252-991A-32227	Sequence 32227, A
21	96.5	8.4	529	3	US-09-247-806-2	Sequence 2, Appli
22	96	8.4	456	4	US-09-252-991A-19417	Sequence 19417, A
23	96	8.4	1958	1	US-07-945-283-2	Sequence 2, Appli
24	95.5	8.3	397	4	US-09-252-991A-25695	Sequence 25695, A
25	95.5	8.3	656	4	US-09-252-991A-22270	Sequence 22270, A
26	94.5	8.2	425	4	US-09-252-991A-31834	Sequence 31834, A
27	94.5	8.2	959	1	US-08-395-580-2	Sequence 2, Appli



QY 81 QOQDVSRCAAIRSMVRYEE--AMPPLEKGMWPGQQPPQ-----GGGGGG 132  
DB 32 QOQDVSRCAAIRSMVRYEE--AMPPLEKGMWPGQQPPQ-----GGGGGG 89  
QY 133 GYYP---CSRPGEGYGGGQGMVPP---PCRPTT---GGGPRIGRVLTKAREVA 182  
DB 90 KRYPSVTCQ-----QVSYPGQASQPSRSSSVHVSVEHQASLKVAKAQA 138  
QY 183 AGLPMCMRLSEBQECIFSGD 204  
DB 139 AOLPAMCRL-----EGGD 151

## RESULT 2

US-09-252-991A-27249  
; Sequence 27249, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; PRIOR FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 27249  
; LENGTH: 308  
; TYPE: PRT  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-27249

Query Match 10.0%; Score 115; DB 4; Length 308;  
Best Local Similarity 25.0%; Pred. No. 0.0012;  
Matches 64; Conservative 13; Mismatches 69; Indels 110; Gaps 14;  
QY 6 AAAAAALCFALVAVACQEVERQRLDLQCVQEVQESPLDACRQVLDRLTGGGGGG 65  
DB 2 SAAARTASGAVATAPAG-----AGPGRGC 28  
QY 66 VGFPRWG-----TGLRM-----RCCQOQDVSRCAAIRSMVRY 102  
DB 29 SGPGRAGRSIPAPRRGPRLARPSRGRTRAPGRCCSGMW--STGSPMAPVRSACRG 86  
QY 103 EAMPPLEKGMW-----WGRQOOP-----PPQGGG-----GGGY 134  
DB 87 SR---PVR--WPGPPAGRAGGYGRRRGRPRRSVPPTVASVAGPSRSGAGPARRAAGPA- 140  
QY 135 YPCSRPGEGYGYG-----QGGQRMVPPC-----RP-GTTGGPGRIGRVLTKARE 180  
DB 141 -APCPGGRSGAARRHRGRRRGRSRRWCPCAGSARRPSGCTGCEPRLPR---RAGRR 196  
QY 181 YAAGLPMCMRLSEBQ 196  
DB 197 AAGRPCCPPGAPGE 212

## RESULT 3

US-09-252-991A-25812  
; Sequence 25812, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; PRIOR FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788

; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 25812  
; LENGTH: 410  
; TYPE: PRT  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-25812

Query Match 10.0%; Score 114.5; DB 4; Length 410;  
Best Local Similarity 36.2%; Pred. No. 0.0018;  
Matches 47; Conservative 10; Mismatches 54; Indels 19; Gaps 8;  
QY 61 GGGGG-----VGFPR-WGTGLRMCCQOQDVSRCAAIRSMVRYEAMPPLEKGMW 115  
DB 81 GAGGRCRRAEPWRTQGTGRRAPAA---VARACRCLPRQRIACVARTAPGERRAG 136  
QY 116 WGRQOQPPPGGGGGGGYVYPCSRPGEYG-YGGGQRMVPPCPRGTTGGPGRIGRVR 174  
DB 137 -GRPR--SAGGAPRGGGAGRLAPPAGPGRGTGQAGRR-----CRPGTGGHR--RPR 187  
QY 175 LTKAREYAAAG 184  
DB 188 PEPRGDGAAG 197

## RESULT 4

US-07-918-023-2  
; Sequence 2, Application US/07918023  
; Patent No. 5427934  
; GENERAL INFORMATION:  
; APPLICANT: Zimmermann, Thomas  
; APPLICANT: Robins, Karen  
; APPLICANT: Birch, Olwen  
; APPLICANT: Bohlen, Elisabeth  
; TITLE OF INVENTION: GENETIC ENGINEERING PROCESS FOR THE  
; TITLE OF INVENTION: PRODUCTION OF S-(+)-2,2-DIMETHYLCYCLOPROPANECARBOXAMIDE  
; TITLE OF INVENTION: BY MICROORGANISMS  
; NUMBER OF SEQUENCES: 2  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Fisher, Christen & Sabol  
; STREET: 2000 M Street, N.W., Suite 590  
; CITY: Washington  
; STATE: DC  
; COUNTRY: USA  
; ZIP: 20036  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07/918,023  
; FILING DATE: 19920724  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: CH 2247/91  
; FILING DATE: 26-JUL-1991  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Marsh, Virgil H.  
; REGISTRATION NUMBER: 23,083  
; REFERENCE/DOCKET NUMBER: LP 1521  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (202) 659-2000  
; TELEFAX: (202) 659-2015  
; TELEX: 248748  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 426 amino acids  
; TYPE: AMINO ACID  
; TOPOLOGY: linear  
; PUBLICATION INFORMATION:

DOCUMENT NUMBER: IE 9224406  
FILING DATE: 23-JUL-1992  
RELEVANT RESIDUES IN SEQ ID NO: 2: FROM 1 TO 426  
PUBLICATION INFORMATION:  
DOCUMENT NUMBER: JP 198717/92  
FILING DATE: 24-JUL-1992  
RELEVANT RESIDUES IN SEQ ID NO: 2: FROM 1 TO 426  
PUBLICATION INFORMATION:  
DOCUMENT NUMBER: PL P 295408  
FILING DATE: 24-JUL-1992  
RELEVANT RESIDUES IN SEQ ID NO: 2: FROM 1 TO 426  
PUBLICATION INFORMATION:  
DOCUMENT NUMBER: RO 92-01033  
FILING DATE: 24-JUL-1992  
RELEVANT RESIDUES IN SEQ ID NO: 2: FROM 1 TO 426  
PUBLICATION INFORMATION:  
DOCUMENT NUMBER: SU UNKNOWN  
FILING DATE: 24-JUL-1992  
RELEVANT RESIDUES IN SEQ ID NO: 2: FROM 1 TO 426  
PUBLICATION INFORMATION:  
DOCUMENT NUMBER: CS PV2323-92  
FILING DATE: 24-JUL-1992  
RELEVANT RESIDUES IN SEQ ID NO: 2: FROM 1 TO 426  
PUBLICATION INFORMATION:  
DOCUMENT NUMBER: HU P9202439  
FILING DATE: 24-JUL-1992  
RELEVANT RESIDUES IN SEQ ID NO: 2: FROM 1 TO 426  
US-07-918-023-2

Query Match 9.5%; Score 109.5; DB 1; Length 426;  
Best Local Similarity 22.8%; Pred. No. 0.006;  
Matches 64; Conservative 16; Mismatches 70; Indels 131; Gaps 12;  
QY 24 QGEVERQRLDLQCWQVQESPLD--ACRQVLDRLTGTGGGGGV----- 66  
DB 123 QAAADRRRLRPSPDRDCHPQPLERPMARGLVQRLGRGHGGAVLRIAGHGRHLPLS 182  
QY 67 -----GPFWRGTGLMRCCQQLQDVSRRCRCRAIRSMVRYGEEAMPPL- 98  
DB 183 IGRQHHGAQHLGQGEPPRLRTGRVPFHR-----PDGAQCRC-----RSH 226  
QY 99 VRGYEAMP-----PLEK-----CWMP-----WGRQQQPPPGGGGGGG 133  
DB 227 ARGHRRGGGAGPYQPVQARLSGHDHARILRPAGHPGAMGTGRRCCLPPGGGAGPGG 286  
QY 134 YYPGSRPGEYGGGGQGMYP--PCRPTTGGG----- 167  
DB 287 -----GAAPGGQAGGFLSRCHPG--GGGLAGAVRGGRRRGARRHVPCTARGL 332  
QY 168 -PRIGRVRLTKAREYAAGLPMCMCR-----LSEPOECISFGG 203  
DB 333 WPARRVDRPGAGPVHRHLPAAAAAPRGLHGPQACTLRAGG 373

RESULT 5  
US-09-252-991A-17887  
; Sequence 17887, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; CURRENT FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 17887  
; LENGTH: 476  
; TYPE: PRT

; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-17887

Query Match 9.0%; Score 103; DB 4; Length 476;  
Best Local Similarity 25.9%; Pred. No. 0.03;  
Matches 60; Conservative 17; Mismatches 79; Indels 76; Gaps 10;  
QY 2 AKIAAAAAALCFAALVAVAVQGE-VERQRLDLQCWQVQESPLDCAQVLDRLTGTG 60  
DB 137 ASRSADAARAVPAIAGASAAERQGRPAGDQRLR-----PADR-----DRLADGR 180  
QY 61 GGGGGVGFWRGTGLMRCCQQLQDVSRRCRCRAIRSMVRYGEEAMPPL-----EKG 112  
DB 181 RGGAG-----QPAGLHLRAWQ-----GRADREVAATAAAGRGGAITARIAAGAAGSADR 231  
QY 113 WWPWGRQQPPPGGGGGGGYYPGSRPGE-----GYGOGG----- 151  
DB 232 -----GAQGARPPQGRAGGQGRTHPGARGDGLARRGEAGLHGYGAGATAGPRLGAG 287  
QY 152 -----QRQWPPCRPTTGGG--PRIGRVRLTKAREYAAGLP 186  
DB 288 LAAARQGLDGLALNLDRLTRPTCRPGSGADVEPAAGLADDTFARVAVRGDP 339

RESULT 6  
US-09-252-991A-19574  
; Sequence 19574, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; CURRENT FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 19574  
; LENGTH: 863  
; TYPE: PRT  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-19574

Query Match 8.8%; Score 101.5; DB 4; Length 863;  
Best Local Similarity 29.5%; Pred. No. 0.082;  
Matches 39; Conservative 14; Mismatches 48; Indels 31; Gaps 7;  
QY 59 GGGGGVGFWRGTGLMRCCQQLQDVSRRCRCRAIRSMVRYGEEAMPPLKGWHPW-- 116  
DB 155 GAAGPCAVGRNRSTGIRRL-----GTRRTAVRPTLRRCRTSVRPVH---LPWLS 202  
QY 117 GRQQPPPGGGGGGGYYPGSRPGEYGGGGQGMYPPCRPTTGGGPRIGRVRLT 176  
DB 203 GR-ADPPYRPHRGQACPAPTGRPGG-----RGERR---PVVRTPGAPRL----- 247  
QY 177 KAREYAAGLPM 188  
DB 248 --RRPAGGVRL 257

RESULT 7  
US-09-252-991A-28410  
; Sequence 28410, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A

```

; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 28410
; LENGTH: 535
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-28410

```

Query Match	8.8%	Score	100.5	DB	4	Length	535
Best Local Similarity	26.7%	Pred.	Ni.0.06				
Matches	63	Conservative	12	Mismatches	71	Indels	87
Gaps	12						
Qy	25	GEVERQLRLDQCQVQESPLDACROVLDRQLTGG-----GGGGGVPGRWGTGL	75				
Db	167	QDRDRPADRAQ-----PAAACR---RHARGNPRGVLGPAPRPPVGEHPWHPGL	214				
Qy	76	RMRCOOQLDYSR-----ECCAAIRSMVRCYEAMPPLEKGVWPPBQ-----	119				
Db	215	RAG-----QHVGRRHAPELPAERPPAAQORDALRESRAA---ORGGHRCQGLGHGRRRG	265				
Qy	120	-----QQPP-----POGGGGGGGGY--YPCSPRG-----EGYGVYGGQGR	153				
Db	266	DRDRQLPHRGPRPGQAGQGRWPGAPHQPPGRRQWHPLRQRGLRHRHRLGHAGRR	325				
Qy	154	QMYP--CRPGTTGGPRIGVRILTKAERYAAGLPMCMRISEPOECISFGG	203				
Db	326	QRAPPGLRLPGDQOHWAARHRLVQSRGAAG-----QAFAGG	364				

```

RESULT 8
US-09-252-991A-30648
; Sequence 30648, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 30648
; LENGTH: 209
; TYPE: prt
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-30648

```

Query Match	8.7%	Score 100;	DB 4;	Length 209;
Best local similarity	27.1%;	Pred. No. 0.023;		
Matches	52;	Conservative	12;	Mismatches 70; Indels 58; Gaps 11;
Qy	48	ACRQVLDRQLTGTGGGGGG-----VGPFWMGTGLRMRCQCOLODVRECRCAAIRSVNR	100	
Db	17	AQRPARHRVAAGGGDGAALPAYLAGAHRAARG--QRAVRLDRPQAORRAGGLHT---	71	
Qy	101	GYEAMP--LEKSWPMGRQQQPPQGG-----GGGGGGVYTPC-----SRP	141	
Db	72	-----WEPGLLEPG--PGRATHDDPAPGHTAGNRAGDPFGYRYGARGPCHRRHRP	124	
Qy	142	GEGY-----GYGGGQRQWYPCPGTGTGGGPR--IGRVV---LTKAREYAA	183	
Db	125	GPQORTAAVDGDCGPRGRGARGAQAQSPAVAPGERRGAQPLVLRSWPTLPDDRQSA	184	
Qy	184	GLPMMCRLESPQ	195	

```

Db          185 VYP-CORSAPE 194

RESULT 9
US-08-343-443B-2
; Sequence 2, Application US/08343443B
; Patent No. 5968734
; GENERAL INFORMATION:
; APPLICANT: Aurtas, Alain
; APPLICANT: Delattre, Olivier
; APPLICANT: Desmaze, Chantal
; APPLICANT: Welot, Thomas
; APPLICANT: Peter, Martine
; APPLICANT: Ploougastel, Beatrice
; APPLICANT: Thomas, Gilles
; APPLICANT: Zuchman, Jessica
; TITLE OF INVENTION: NUCLEIC ACID
; TITLE OF INVENTION: CHROMOSOME 2
; TITLE OF INVENTION: TRANSLATIONS
; TITLE OF INVENTION: TUMORS, AND
; TITLE OF INVENTION: TRANSLOCATION
; NUMBER OF SEQUENCES: 129
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Weiser & Associates
; STREET: 230 South Fifteenth Street
; CITY: Philadelphia
; STATE: PA
; COUNTRY: USA
; ZIP: 19102
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: AEDIT 1.0 DOS text editor
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/343,
; FILING DATE: 18-NOV-1994
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/FR93/0/30
; FILING DATE: 19-MAY-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: FR 92/0612
; FILING DATE: 20-MAY-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Weiser, Gerard J.
; REGISTRATION NUMBER: 19,763
; REFERENCE/DOCKET NUMBER: 989.6
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-875-8393
; TELEFAX: 215-875-8394
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 656 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-143-443B-2

```

```

Query Match      8.7%; Score 100; DB 2; Length 656;
Best Local Similarity 38.4%; Pred. No. 0.084;
Matches 28; Conservative 6; Mismatches 25; Indels 14; Gaps 3;

Qy    99 VRGYEEAMPLEKGNWPNWGROQQPPPOGGGGGQGGYYYPCSRPGEGYGYGGGGORQMYP 158
      ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db    454 MRG---GLPRE-----GRGMPPPLRGFGPGCGPGPMGRMG-----GRGGDRGGFP 499

Qy    159 CRFGTTGGGPRI 171
      :|||::|::|
Db    500 RGPRGRSGNPSGG 512

```

US-09-214-564A-4  
; Sequence 4, Application US/09214564A  
; Patent No. 6150515  
; GENERAL INFORMATION:  
; APPLICANT: Sharp, Phillip A.  
; APPLICANT: Zhou, Qiang  
; TITLE OF INVENTION: Tat-SF: Cofactor For Stimulation Of Transcriptional  
; FILE REFERENCE: M0656/7042  
; CURRENT APPLICATION NUMBER: US/09/214,564A  
; PRIOR FILING DATE: 1999-08-18  
; PRIOR FILING DATE: 1996-07-03  
; PRIOR APPLICATION NUMBER: US 60/033,152  
; PRIOR FILING DATE: 1996-12-13  
; PRIOR APPLICATION NUMBER: PCT/US97/11713  
; PRIOR FILING DATE: 1997-07-03  
; NUMBER OF SEQ ID NOS: 5  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 4  
; LENGTH: 656  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-214-564A-4

Query Match 8.7%; Score 100; DB 3; Length 656;  
Best Local Similarity 38.4%; Pred. No. 0.084;  
Matches 28; Conservative 6; Mismatches 25; Indels 14; Gaps 3;

QY 99 VRGYEAMPLEKGNWPMGRQPPQGGGGGGYYPGCRPGEYGYGGGQGMYP 158  
DB 454 MRG---GLPPRE-----GRGMPPPLRGPGGPGPMGRMG-----GRGGDRGGFPF 499

QY 159 CRPGTGGGPRIG 171  
DB 500 RGRGRSGRGNPSG 512

RESULT 11  
US-09-252-991A-17915  
; Sequence 17915, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; PRIOR FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 17915  
; LENGTH: 238  
; TYPE: PRT  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-17915

Query Match 8.6%; Score 99; DB 4; Length 238;  
Best Local Similarity 28.1%; Pred. No. 0.033;  
Matches 39; Conservative 8; Mismatches 56; Indels 36; Gaps 7;

QY 49 CROVLDRQLTG-----GGGGGGVPRFMGTGLRMRCQQQLQDVSRRCRAIRSMV 102  
DB 114 CRTVARRRATGRRGASSGGAAGSAPRATAPGIRRR-----SVPARCGCVA-----AGY 162

QY 103 EAMPLEKGNWPMGRQPP-----PQGGGGGGYYPGCRPGEYGYGGGQGMYP 156  
DB 163 WRSAREPLA---WPRGRGSAFVRGPRRRSGSAPGTAPRRRTGTGRAIRHDSGRHRAVC 219

QY 157 PP-----CRPGTTG 165

DB 220 VPASACGARCGAARPGACG 238

RESULT 12  
US-08-216-846-2  
; Sequence 2, Application US/08216846  
; Patent No. 5444046  
; GENERAL INFORMATION:  
; APPLICANT: MIYAZAKI, TOSHIYUKI  
; APPLICANT: MORIMOTO, TOSHIHISA  
; APPLICANT: MURAYAMA, RYOJI  
; APPLICANT: MATSUBARA, HIROSHI  
; TITLE OF INVENTION: AMYLASE INHIBITORS  
; NUMBER OF SEQUENCES: 4  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MATER & NEUSTADT,  
; ADDRESS: P.C.  
; STREET: 1755 S. Jefferson Davis Highway, Suite 400  
; CITY: Arlington  
; STATE: Virginia  
; COUNTRY: U.S.A.  
; ZIP: 22202  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/216,846  
; FILING DATE: 24-MAR-1994  
; CLASSIFICATION: 530  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: JP 91881/1993  
; FILING DATE: 29-MAR-1993  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: JP 148423/1993  
; FILING DATE: 28-MAY-1993  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Oblon, No. 5444046man F.  
; REGISTRATION NUMBER: 24,618  
; REFERENCE/DOCKET NUMBER: 1327-029-0X  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (703) 413-3000  
; TELEFAX: (703) 413-2220  
; TELEX: 248855 OPAT UR  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 124 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-216-846-2

Query Match 8.6%; Score 98.5; DB 1; Length 124;  
Best Local Similarity 27.6%; Pred. No. 0.018;  
Matches 43; Conservative 8; Mismatches 38; Indels 67; Gaps 7;

QY 39 QEVQESPLDACRQVLDRLQLTGGGGGGVGFPRMGTGLRMRCQQQLQDVSRRCRAIRSM 98  
DB 10 QAFQVPALPACRPLRLRLQCNQSGVPEAV-----LR-DCCQLAHISEWCRCGALYSM 60

QY 99 VRG-YEEAMPLEKGNWPMGRQPPQGGGGGGYYPGCRPGEYGYGGGQGMYP 157  
DB 61 LDSMYKE-----HGAQEGQAG----- 76

QY 158 PCRPGTTGGPRIGR--VELTKAREYAAGLPMWCL 191  
DB 77 -----TGAPRCRRREVVKLT-----AASITAVCRL 101

RESULT 13  
US-08-782-177-3

Qy	25	GEVEROR--LRDL--	--QCWQEVQESPLDACROVLDRLTGGG--	-----GGVGPPFW	71
Db	117	GLVEHORAGLRRLARRHVLGGQRHPLPALHRAGR--	--GGAGVHPQAAEGTGLHPVR	173	
Qy	72	GTGLRW-----	--RCCQQLQDVSRRCRAAIRSMVR-----	-----	100
Db	174	ARGHRVGDPCAGDHRCAPHAGPAINESHAGLRHLRLQRRRRGDRGCRPTARRYPAA	233		
Qy	101	GVEEAMPP--LEKGWPMWGQQPPPOGG--GGGQGGYY--	YPCSRFEGYGVGGGQOROM	155	
Db	234	GTGNNPLPQWRDRGVWFLRGAGPPGQRRRLAQOAGALRRLPGGRFGD-----	PRLL	286	
Qy	156	YPPCRFGTTGGGPRIGRV--RLTKAREYAAGLPMPCR	190		

```

; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
;

```

Db 287 HPP-----GGGRAGGHLPRRLPRARQHL--LPFRLR 315  
: || | | : | | | | : | | | |

Search completed: November 29, 2003, 17:07:34  
Job time : 25 secs

**THIS PAGE BLANK (USPTO)**



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 29, 2003, 17:06:40 ; Search time 46 Seconds  
(without alignments)  
825.985 Million cell updates/sec

Title: US-10-053-410-4

Perfect score: 1148

Sequence: 1 MAKIAAAAAALCFALNAV.....MMCRILSEPOECIFSGGDQY 206

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 673684 seqs, 184443283 residues

Total number of hits satisfying chosen parameters: 673684

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications\_AA:\*

- 1: /cgn2\_6/ptodata/2/pubpaa/US07\_PUBCOMB.pcp.\*
- 2: /cgn2\_6/ptodata/2/pubpaa/PCT\_NEW\_PUB.pcp.\*
- 3: /cgn2\_6/ptodata/2/pubpaa/US06\_NEW\_PUB.pcp.\*
- 4: /cgn2\_6/ptodata/2/pubpaa/US06\_PUBCOMB.pcp.\*
- 5: /cgn2\_6/ptodata/2/pubpaa/US07\_NEW\_PUB.pcp.\*
- 6: /cgn2\_6/ptodata/2/pubpaa/PCTUS\_PUBCOMB.pcp.\*
- 7: /cgn2\_6/ptodata/2/pubpaa/US08\_NEW\_PUB.pcp.\*
- 8: /cgn2\_6/ptodata/2/pubpaa/US08\_PUBCOMB.pcp.\*
- 9: /cgn2\_6/ptodata/2/pubpaa/US09\_PUBCOMB.pcp.\*
- 10: /cgn2\_6/ptodata/2/pubpaa/US09\_PUBCOMB.pcp.\*
- 11: /cgn2\_6/ptodata/2/pubpaa/US09C\_PUBCOMB.pcp.\*
- 12: /cgn2\_6/ptodata/2/pubpaa/US09\_NEW\_PUB.pcp.\*
- 13: /cgn2\_6/ptodata/2/pubpaa/US10A\_PUBCOMB.pcp.\*
- 14: /cgn2\_6/ptodata/2/pubpaa/US10B\_PUBCOMB.pcp.\*
- 15: /cgn2\_6/ptodata/2/pubpaa/US10C\_PUBCOMB.pcp.\*
- 16: /cgn2\_6/ptodata/2/pubpaa/US10\_NEW\_PUB.pcp.\*
- 17: /cgn2\_6/ptodata/2/pubpaa/US60\_NEW\_PUB.pcp.\*
- 18: /cgn2\_6/ptodata/2/pubpaa/US60\_PUBCOMB.pcp.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	117	10.2	169	12	US-10-165-289A-4
2	106.5	9.3	1145	10	US-09-866-557A-5
3	106.5	9.3	1145	15	US-10-055-797-5
4	105	9.1	168	12	US-10-165-289A-2
5	105	9.1	277	11	US-09-847-102A-73
6	105	9.1	694	11	US-09-847-102A-56
7	105	9.1	694	12	US-10-205-976-53
8	105	9.1	694	15	US-10-205-823-144
9	101	8.8	591	15	US-10-233-553-23
10	101	8.8	894	15	US-10-233-553-11
11	100.5	8.8	544	12	US-10-017-161-670
12	97.5	8.5	302	14	US-10-029-180-18
13	97.5	8.5	623	15	US-10-156-761-11120
14	96.5	8.4	529	9	US-09-861-597-2
15	96	8.4	2331	15	US-10-184-644-59

16	96	8.4	2331	15	US-10-184-634-59	Sequence 59, Appl
17	94.5	8.2	520	12	US-10-168-097A-36	Sequence 36, Appl
18	94.5	8.2	520	12	US-10-239-431A-34	Sequence 34, Appl
19	94.5	8.2	859	15	US-10-153-668-324	Sequence 324, Appl
20	94.5	8.2	892	15	US-10-153-668-438	Sequence 438, Appl
21	94	8.2	485	10	US-09-925-300-1664	Sequence 1664, Ap
22	94	8.2	775	15	US-10-156-761-12824	Sequence 12824, A
23	93.5	8.1	132	9	US-09-864-761-43644	Sequence 43644, A
24	93.5	8.1	166	15	US-10-157-031-80	Sequence 80, Appl
25	93.5	8.1	238	15	US-10-204-887-116	Sequence 116, Appl
26	92.5	8.1	599	11	US-09-847-102A-37	Sequence 37, Appl
27	92.5	8.1	755	10	US-09-919-497-57	Sequence 57, Appl
28	92	8.0	274	11	US-09-847-102A-74	Sequence 74, Appl
29	92	8.0	682	15	US-10-152-548-16	Sequence 16, Appl
30	92	8.0	685	11	US-09-847-102A-57	Sequence 57, Appl
31	92	8.0	685	15	US-10-146-474-14	Sequence 14, Appl
32	91.5	8.0	148	12	US-10-447-541-6	Sequence 6, Appl
33	91.5	8.0	253	12	US-10-304-630-7	Sequence 7, Appl
34	91.5	8.0	253	12	US-10-304-630-9	Sequence 9, Appl
35	91.5	8.0	253	12	US-10-304-630-10	Sequence 10, Appl
36	91.5	8.0	253	12	US-10-304-630-11	Sequence 11, Appl
37	91.5	8.0	253	12	US-10-304-630-12	Sequence 12, Appl
38	91.5	8.0	253	12	US-10-304-630-14	Sequence 14, Appl
39	90.5	7.9	276	12	US-10-017-161-2268	Sequence 2268, Ap
40	90.5	7.9	481	15	US-10-156-761-12906	Sequence 12906, A
41	90.5	7.9	752	15	US-10-156-761-9350	Sequence 9350, Ap
42	90.5	7.9	1058	15	US-10-174-363-38	Sequence 38, Appl
43	90.5	7.9	3060	15	US-10-184-644-337	Sequence 337, Appl
44	90.5	7.9	3060	15	US-10-184-634-337	Sequence 337, App
45	90	7.8	378	9	US-09-849-967A-2	Sequence 2, Appl

ALIGNMENTS

RESULT 1

US-10-165-289A-4  
; Sequence 4, Application US/10165289A  
; Publication No. US2003015917A1  
; GENERAL INFORMATION:  
; APPLICANT: QIU, Xiao  
; APPLICANT: TRUSKA, Martin  
; APPLICANT: HU, Zhiyuan  
; TITLE OF INVENTION: Flax (linum usitatissimum L.)  
; TITLE OF INVENTION: Seed-Specific Promoters  
; FILE REFERENCE: BNZ-005  
; CURRENT APPLICATION NUMBER: US/10/165,289A  
; PRIOR FILING DATE: 2002-06-06  
; PRIOR APPLICATION NUMBER: 60/295823  
; NUMBER OF SEQ ID NOS: 13  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 4  
; LENGTH: 169  
; TYPE: PRT  
; ORGANISM: Linum usitatissimum  
US-10-165-289A-4

Query Match	10.2%	Score 117;	DB 12;	Length 169;
Best Local Similarity	23.9%	Pred. No. 0.0097;		
Matches	56;	Conservative 24;	Mismatches 56;	Indels 98; Gaps 13;
QY	1	MAKIAAAAAALCFALNAVAVACQEVERQRRLDQCQVEQSPLDACRQ----	VLDROL 57	
DB	1	MAKIASLAATAFLFLIVV-----	DASVTTVTIIDE- 33	
QY	58	TGGGGGGGGVGPFRWGTGLMRCCQOL--	QDVSRRCRCALRSMVRGVEAMPLEKWWP 115	
DB	34	TWGGGGGGG-----	GCGQQQSCBQIQQDFLRSCQ-----	QFM 68
QY	116	WGRQQQPPPGGGGGGGGGYYPFCRPGEGYGGGQGRQMYPPC-----	RPGTGGG- 167	
DB	69	WEKVQR-----	GGRSHYY-----	NQGRGGGGSQYFDSCCDDLQLSTGCTCRGL 113

```
QY 168 -PRIGVR-----LTKAREYAAGLPMWCRSLRQECISIFSGDQ 205
Db 114 ERAIGMRQEIQQGQGVQVORWIOQAQTAKDLPGQCR-TQPSQCG-FQGGQQ 165

RESULT 2
US-09-866-557A-5
; Sequence 5, Application US/09866557A
; Patent No. US20020162126A1
; GENERAL INFORMATION:
; APPLICANT: Hammond, S.
; APPLICANT: Hammond, G.
; APPLICANT: Beach, D.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR RNA INTERFERENCE
; FILE REFERENCE: GNCA-P03-007
; CURRENT APPLICATION NUMBER: US/09/866,557A
; CURRENT FILING DATE: 2002-03-26
; PRIOR APPLICATION NUMBER: 60/189,739
; PRIOR FILING DATE: 2000-03-16
; PRIOR APPLICATION NUMBER: 60/243,097
; PRIOR FILING DATE: 2000-10-24
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 5
; LENGTH: 1145
; TYPE: PRT
; ORGANISM: Drosophila melanogaster
US-09-866-557A-5

Query Match 9.3%; Score 106.5; DB 10; Length 1145;
Best Local Similarity 25.9%; Pred. No. 0.54;
Matches 45; Conservative 25; Mismatches 71; Indels 33; Gaps 6;

QY 30 QRLRLQWCQEVQESPLDACRQVLDRLTGGGGGGVGPFRWG-----TGLRM 77
Db 47 QQLQPPQ--QQQQQPHQQQQSSRQPSSTSSGSRASGFQQGGQQQKSDAEGWTAQKK 104
QY 78 RCCQQLQDVSRRCRA-----AIRSMVRGYEEAMPPEKGMWPGRQ-----QQPPP--- 124
Db 105 QGKQVQGWTKQGQQGGHQQGQDGGYQQRPPGQQGGHQQGQGGYQQRPPGQQ 164
QY 125 -----QGGGGGGGGYVY-PCSRPGEYGYGGGQROMYPPCRPTGTTGGGPRIGR 172
Db 165 QGGHQQGQGGYQQRPPSGQQGGHQQGQGGYQQRPPGQQGGHQQGR 218

RESULT 3
US-10-055-797-5
; Sequence 5, Application US/10055797
; Publication No. US2003008471A1
; GENERAL INFORMATION:
; APPLICANT: Hammond et al.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR RNA INTERFERENCE
; FILE REFERENCE: GNCA-P03-007
; CURRENT APPLICATION NUMBER: US/10/055,797
; CURRENT FILING DATE: 2002-01-22
; PRIOR APPLICATION NUMBER: 60/189,739
; PRIOR FILING DATE: PCT/US01/08435
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 5
; LENGTH: 1145
; TYPE: PRT
; ORGANISM: Drosophila melanogaster
US-10-055-797-5

Query Match 9.3%; Score 106.5; DB 15; Length 1145;
Best Local Similarity 25.9%; Pred. No. 0.54;
Matches 45; Conservative 25; Mismatches 71; Indels 33; Gaps 6;

QY 30 QRLRLQWCQEVQESPLDACRQVLDRLTGGGGGGVGPFRWG-----TGLRM 77
```

```
Db 47 QQLQPPQ--QQQQQPHQQQQSSRQPSSTSSGSRASGFQQGGQQQKSDAEGWTAQKK 104
QY 78 RCCQQLQDVSRRCRA-----AIRSMVRGYEEAMPPEKGMWPGRQ-----QQPPP--- 124
Db 105 QGKQVQGWTKQGQQGGHQQGQDGGYQQRPPGQQGGHQQGQGGYQQRPPGQQ 164
QY 125 -----QGGGGGGGGYVY-PCSRPGEYGYGGGQROMYPPCRPTGTTGGGPRIGR 172
Db 165 QGGHQQGQGGYQQRPPSGQQGGHQQGQGGYQQRPPGQQGGHQQGR 218

RESULT 4
US-10-165-289A-2
; Sequence 2, Application US/10165289A
; Publication No. US20030159174A1
; GENERAL INFORMATION:
; APPLICANT: QIU, Xiao
; APPLICANT: TRUSKA, Martin
; APPLICANT: HU, Zhiyuan
; TITLE OF INVENTION: Flax (Linum usitatissimum L.)
; FILE REFERENCE: BNZ-005
; CURRENT APPLICATION NUMBER: US/10/165,289A
; CURRENT FILING DATE: 2002-06-06
; PRIOR APPLICATION NUMBER: 60/295823
; PRIOR FILING DATE: 2001-06-06
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 168
; TYPE: PRT
; ORGANISM: Linum usitatissimum
US-10-165-289A-2

Query Match 9.1%; Score 105; DB 12; Length 168;
Best Local Similarity 26.5%; Pred. No. 0.11;
Matches 39; Conservative 21; Mismatches 45; Indels 42; Gaps 7;

QY 1 MAKIAAAAAALCFPAALVAVVC-----QGEVERQRLRLDLCWQEVQ 43
Db 1 MAKLAALAAVATAFLFLIVVDASVTRTVIDEDTNGRGQGGGQQGQ-----QCEKIQIE 56
QY 44 SP-LDACROVL-DRLTGG-----GGGGGGVGPFRWGTLRMRCQQLQDVSRRCRA 94
Db 57 QDYLRCQQLFLWEKVQKGRSYNYNGRGQGGGQQGQ-----CCDLKQLRSECTC-- 109
QY 95 IRSMVRGYEEAMPPEKGMWPGRQ 121
Db 110 -----RGLERAIGMRQDIQQGGQQQE 131

RESULT 5
US-09-847-102A-73
; Sequence 73, Application US/09847102A
; Publication No. US2003004409A1
; GENERAL INFORMATION:
; APPLICANT: University of California
; APPLICANT: Carson, Dennis A.
; APPLICANT: Carr, Maribat
; APPLICANT: Rhee, Chae-Seo
; APPLICANT: Lorenzo, Leoni M.
; APPLICANT: Malini, Sen
; TITLE OF INVENTION: IMMUNOLOGIC COMPOSITIONS AND METHODS FOR
; FILE REFERENCE: 22000-20629.00
; CURRENT APPLICATION NUMBER: US/09/847,102A
; CURRENT FILING DATE: 2001-05-01
; NUMBER OF SEQ ID NOS: 138
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 73
; LENGTH: 277
; TYPE: PRT
; ORGANISM: Homo sapiens
```



```

; PRIOR APPLICATION NUMBER: 60/325,020
; PRIOR FILING DATE: 2001-09-25
; PRIOR APPLICATION NUMBER: 60/341,746
; PRIOR FILING DATE: 2001-12-12
; PRIOR APPLICATION NUMBER: 60/362,158
; PRIOR FILING DATE: 2002-03-05
; NUMBER OF SEQ ID NOS: 455
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 144
; LENGTH: 694
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-205-823-144

Query Match
Best Local Similarity 9.1%; Score 105; DB 15; Length 694;
Matches 51; Conservative 22; Mismatches 72; Indels 70; Gaps 10;

QY 5 AAAAAAALCFAAVLAVAVACOG-----EVER-QRLRLDQCQEV 41
DB 25 AAAAASAKELACQETVPLCKIGVNYTYMPNQFNHDTQDEAGLEVHQFWPLVEIQSPDL 84

QY 42 Q-----ESPLDACRQVLDRLQLTGGGGGGVGP-----FRWGTGLMRCC 80
DB 85 KFFLCSMYTPICLEDYKKPLPPCRSVCE-----AKAGCAPLMRQYGFAPWD--NRCD 136

QY 81 QQLQDVSRERC-----AATRSVMRGYEEAAMPLEKGNWPNWGRQQPPPOGGGGGGGY 135
DB 137 RLPEQGNPTLCMDYNRDTLTAAPSPPRLPPPPG-----EQPPSGSGHGRP--- 185

QY 136 YPCSR-PGEGYGYGGGQGMYPPCRGTGGGPR 169
DB 186 -PGARPPHRRGGGRGGGGGDAAPARGGGGGKAR 219

RESULT 9
US-10-233-553-23
; Sequence 23, Application US/10233553
; Publication No. US20030125285A1
; GENERAL INFORMATION:
; APPLICANT: NIPPON SHINYAKU CO., LTD.
; APPLICANT: HIRABAYASHI, Kazuko
; APPLICANT: YANO, Junichi
; TITLE OF INVENTION: Method of examining the efficacy of therapy with nucleic acid
; FILE REFERENCE: B-345
; CURRENT APPLICATION NUMBER: US/10/233,553
; CURRENT FILING DATE: 2002-12-19
; PRIOR APPLICATION NUMBER: JP 2001-267385
; PRIOR FILING DATE: 2001-09-04
; NUMBER OF SEQ ID NOS: 31
; SEQ ID NO 23
; LENGTH: 591
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-233-553-23

Query Match
Best Local Similarity 8.8%; Score 101; DB 15; Length 591;
Matches 43; Conservative 5; Mismatches 35; Indels 84; Gaps 8;

QY 58 TGGGGGGGVPFRWGTGLMRCCQQLQDVSRRCRCAAIRSMVRG--YEEAMPLEKGV-WW 114
DB 397 SGGGGGGGGSSGYG-----SYQGDNYNSVPPPKHAGKKQ 432

QY 115 PWGRQOQP-----PPOG-----GG 128
DB 433 PHGGQKQPSYGSYQSHQGGQSYNSQSYNSYNSYNSPGG 492

QY 129 GGGGGYYPC-----SRPGEYGYGGGQGMYPPCRGTGGG 167
DB 493 GGGSDNYESKFNSYSGGRSG--GNSYSGG--ASYNPGSHGGYGGG 536

RESULT 10
US-10-233-553-11
; Sequence 11, Application US/10233553
; Publication No. US20030125285A1
; GENERAL INFORMATION:
; APPLICANT: NIPPON SHINYAKU CO., LTD.
; APPLICANT: HIRABAYASHI, Kazuko
; APPLICANT: YANO, Junichi
; TITLE OF INVENTION: Method of examining the efficacy of therapy with nucleic acid
; FILE REFERENCE: B-345
; CURRENT APPLICATION NUMBER: US/10/233,553
; CURRENT FILING DATE: 2002-12-19
; PRIOR APPLICATION NUMBER: JP 2001-267385
; PRIOR FILING DATE: 2001-09-04
; NUMBER OF SEQ ID NOS: 31
; SEQ ID NO 11
; LENGTH: 894
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-233-553-11

Query Match
Best Local Similarity 8.8%; Score 101; DB 15; Length 894;
Matches 43; Conservative 5; Mismatches 35; Indels 84; Gaps 8;

QY 58 TGGGGGGGVPFRWGTGLMRCCQQLQDVSRRCRCAAIRSMVRG--YEEAMPLEKGV-WW 114
DB 700 SGGGGGGGGSSGYG-----SYQGDNYNSVPPPKHAGKKQ 735

QY 115 PWGRQOQP-----PPOG-----GG 128
DB 736 PHGGQKQPSYGSYQSHQGGQSYNSQSYNSYNSYNSPGG 795

QY 129 GGGGGYYPC-----SRPGEYGYGGGQGMYPPCRGTGGG 167
DB 796 GGGSDNYESKFNSYSGGRSG--GNSYSGG--ASYNPGSHGGYGGG 839

RESULT 11
US-10-017-161-670
; Sequence 670, Application US/10017161
; Publication No. US20030143668A1
; GENERAL INFORMATION:
; APPLICANT: SUWA, MAKIKO
; APPLICANT: ASAI, KIYOSHI
; APPLICANT: AKIYAMA, YUTAKA
; APPLICANT: ABURATANI, HIROYUKI
; TITLE OF INVENTION: NOVEL G PROTEIN-COUPLED RECEPTORS
; FILE REFERENCE: 084335/0152
; CURRENT APPLICATION NUMBER: US/10/017,161
; CURRENT FILING DATE: 2002-12-18
; PRIOR APPLICATION NUMBER: JP 2001/246789
; PRIOR FILING DATE: 2001-06-18
; NUMBER OF SEQ ID NOS: 2430
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 670
; LENGTH: 544
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: MOD RES
; LOCATION: (402)...(435)
; OTHER INFORMATION: Variable amino acid
US-10-017-161-670

Query Match
Best Local Similarity 8.8%; Score 100.5; DB 12; Length 544;
Matches 51; Conservative 14; Mismatches 48; Indels 89; Gaps 10;

QY 25 GEVERORLRDLQCVQES--PLDACRQVLDRLQLTGGGGGGVGP-----68
DB 358 GSEHLRLAAAPSGREVGAAPRPASRPPIRE-----GGGGVSPPPGPPPPXXXXXX 410
```

QY 69 -----FRWGTGLRMRCQQQLQDVSRRCRCRAAIRSMVRGYEEA 105  
Db 411 XXXXXXXXXXXXXXXXXXXXWGGQPPSRASRPV-----REVGRASARPLLGSSEP 466  
QY 106 MPELEKRWPGWGRQQQPPGGGGG-----QGGYY-----YPCSRPGGYGYG 148  
Db 467 LCPASR-----PVQEGGGVSPPTGPPQPRQGRGWGVSPPPPGQP-SRPGGRWG-G 513  
QY 149 Q-----GGQROWPPCRP 161  
Db 514 QPPPGQPHPGGEGRLYPASRP 535

RESULT 12  
US-10-029-180-18  
; Sequence 18, Application US/10029180  
; Publication No. US20020182708A1  
; GENERAL INFORMATION:  
; APPLICANT: Cali, Brian M.  
; APPLICANT: Holtzman, Doug  
; APPLICANT: Madden, Kevin T.  
; APPLICANT: Milna, G. Todd  
; APPLICANT: Sherman, Amir  
; APPLICANT: Silva, Jeffrey C.  
; APPLICANT: Trueheart, Josh  
; APPLICANT: Zhang, Lixin  
; TITLE OF INVENTION: No. US20020182708A1e1 Regulators of Fungal Gene Expression  
; FILE REFERENCE: MIC-004  
; CURRENT APPLICATION NUMBER: US/10/029,180  
; CURRENT FILING DATE: 2001-12-22  
; PRIOR APPLICATION NUMBER: US 60/257,431  
; PRIOR FILING DATE: 2000-12-22  
; NUMBER OF SEQ ID NOS: 138  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 18  
; LENGTH: 302  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: fungal gene  
US-10-029-180-18

Query Match 8.5%; Score 97.5; DB 14; Length 302;  
Best Local Similarity 25.7%; Pred. No. 0.87;  
Matches 49; Conservative 16; Mismatches 57; Indels 69; Gaps 10;  
QY 26 EVERQL-----RLQCHQEVQESPLDACRQVLDRLTGG-----GGG 63  
Db 133 EYEQDEMSRANPPVHLTRELCW-AKMLGSDISSYDTKDERGSGPSGAPPPGPPQYGY 191  
QY 64 GGVGPRWGTGLRMRCQQQLQDVSRRCRAAIRSMVRGYEEAMPPLEKRWPGWGRQQQPP 123  
Db 192 GSGP-GYSSAPYNGQ-----QYSPAYPHSQA--P-GYVAPP 230  
QY 124 PGGGGG-----QGGY-----YPCSRPGGYGYGGQGRQMTYPPCRP 161  
Db 231 POGYGYGYSGPGYSGPAPYSSGQGGYGSAPYPPSSQASGYGQQYQGR---PGSQP 287  
QY 162 GTTGGPRGR 172  
Db 288 GYPGQQPPYQ 298

RESULT 13  
US-10-156-761-11120  
; Sequence 11120, Application US/10156761  
; Publication No. US20030119018A1  
; GENERAL INFORMATION:  
; APPLICANT: OMURA, SATOSHI  
; APPLICANT: IKEDA, HARUO  
; APPLICANT: ISHIKAWA, JUN  
; APPLICANT: HORIKAWA, HIROSHI  
; APPLICANT: SHIBA, TADAYOSHI

; APPLICANT: SAKAKI, YOSHIYUKI  
; APPLICANT: HATTORI, MASAHIRA  
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
; FILE REFERENCE: 249-262  
; CURRENT APPLICATION NUMBER: US/10/156,761  
; CURRENT FILING DATE: 2002-05-29  
; PRIOR APPLICATION NUMBER: JP 2001-204089  
; PRIOR FILING DATE: 2001-05-30  
; PRIOR APPLICATION NUMBER: JP 2001-272697  
; PRIOR FILING DATE: 2001-08-02  
; NUMBER OF SEQ ID NOS: 15109  
; SEQ ID NO 11120  
; LENGTH: 623  
; TYPE: PRT  
; ORGANISM: Streptomyces avermitilis  
US-10-156-761-11120

Query Match 8.5%; Score 97.5; DB 15; Length 623;  
Best Local Similarity 38.1%; Pred. No. 1.8;  
Matches 37; Conservative 6; Mismatches 33; Indels 21; Gaps 7;  
QY 102 YEEAMPPLEKGW-WPGRQQQPPP---QGG-GGGGGYYPGSRPGSGYGGCGGQROM 155  
Db 76 YGQPPGQGGQGGPGYGPQPPGYAQQPPGYAPQPGYGPQQP--GYGYPQ-----127  
QY 156 YPPCRPPT---TGGG---PRIGRVRLTKAREYAAAGL 185  
Db 128 QPPQPPGAPTTPPGGGGSKNPKGRPAVILAAVAAL 164

RESULT 14  
US-09-861-597-2  
; Sequence 2, Application US/09861597  
; Patent No. US20020064539A1  
; GENERAL INFORMATION:  
; APPLICANT: PHILLIPPE, Michel  
; APPLICANT: GARSON, Jean-Claude  
; APPLICANT: ARRAUDEAU, Jean-Pierre  
; TITLE OF INVENTION: COSMETIC OR DERMATOLOGICAL COMPOSITION CONTACTING AT  
; TITLE OF INVENTION: LEAST ONE NATURAL OR RECOMBINANT SPIDER SILK OR AN  
; TITLE OF INVENTION: ANALOG  
; FILE REFERENCE: 6388-0365-0  
; CURRENT APPLICATION NUMBER: US/09/861,597  
; CURRENT FILING DATE: 2001-05-22  
; PRIOR APPLICATION NUMBER: 09/247,806  
; PRIOR FILING DATE: 1999-02-11  
; PRIOR APPLICATION NUMBER: FR 98/01614  
; PRIOR FILING DATE: 1998-02-11  
; NUMBER OF SEQ ID NOS: 14  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 2  
; LENGTH: 529  
; TYPE: PRT  
; ORGANISM: Nephila clavipes  
US-09-861-597-2

Query Match 8.4%; Score 96.5; DB 9; Length 529;  
Best Local Similarity 32.8%; Pred. No. 1.9;  
Matches 41; Conservative 9; Mismatches 54; Indels 21; Gaps 5;  
QY 55 RQLTGGGGGGVGPFRWGTGLRMRCQQQLQDVSRRCRAAIRSMVRGYEEAMPPLEKGMW 114  
Db 45 QGPGGYGPGQGGPGRYGFG-----QQPGSGGAAAAAAGSGGQG--PGYGPQQGGP 97  
QY 115 PWGRQQQPPPGG-----GGGGGGYYPGSRPGGYGGCGGQROMYPPCRP 162  
Db 98 GYGQGGQGGSGPGSAAAAAASAAESGGGPGYGPQGGPG--GYGPGQGGPGG-YGPGQGG 155  
QY 163 TTGGG 167  
Db 156 PSGPG 160

```
RESULT 15
US-10-184-644-59
; Sequence 59, Application US/10184644
; Publication No. US20030044930A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Chen, Jian
; APPLICANT: Desnoyers, Luc
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Pan, James
; APPLICANT: Smith, Victoria
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE OF INVENTION: ACIDS ENCODING THE SAME
; FILE REFERENCE: P3430R1C227
; CURRENT APPLICATION NUMBER: US/10/184,644
; CURRENT FILING DATE: 2002-06-28
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 59
; LENGTH: 2331
; TYPE: DNA
; ORGANISM: Homo Sapien
US-10-184-644-59

Query Match      8.4%; Score 96; DB 15; Length 2331;
Best Local Similarity 25.6%; Pred. No. 9.1;
Matches 42; Conservative 2; Mismatches 56; Indels 64; Gaps 7;

Qy      5 AAAAAALCPAALVAVAVCGEVEVERLRDLQCNQEVQESPLDACRQVLDRLQLTGGGGG 64
      |||||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db      2080 AAAAAAACAAACAAACAAACAA-----AAAAAGAC-----ATGGGAGA 2117

Qy      65 GVGPFRRGTG-LRMRCQQLQDVSRRCRAIRSMVRYEEMPPLEKGMWPGRQQQPP 123
      |||||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db      2118 GA---TGGTGTGTTTTC-----TCTCCACCACC----- 2142

Qy      124 PGGGGGGGGYYPYPCSRPGEYGYGGGQRMYPFCRPGTTGGG 167
      |||||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db      2143 -TGGGGATA-----CGATGAGAAGATGCCACCTGCAAGCCAGG 2180
```

Search completed: November 29, 2003, 17:11:45

Job time : 49 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: November 29, 2003, 17:07:10 ; Search time 271 Seconds  
(without alignments)  
2051.971 Million cell updates/sec

Title: US-10-053-410-4

Perfect score: 1148

Sequence: 1 MAKIAAAAAALCFALVAV.....MMCRLEPQECIFSGGDQY 206

Scoring table:

BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 2552756 segs, 1349719017 residues

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame\_p2n.model -DRV=xlh  
-Q=/cgn2\_1/USPTO.spool/US10053410/runat\_28112003\_140756\_27874/app\_query.fasta\_1.391  
-DB=N Geneseq 19Jun03 -QFMT=fastap -SUFFIX=p2n.rng -MINMATCH=0.1 -LOOPCL=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi  
-LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US10053410@cgn\_1\_1\_0 @runat\_28112003\_140756\_27874 -NCPU=6 -ICPU=3  
-NO MMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DSV TIMEOUT=120 -WARN TIMEOUT=30 -THREDS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

N Geneseq 19Jun03:\*

- 1: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT:\*
- 2: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT:\*
- 3: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT:\*
- 4: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT:\*
- 5: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT:\*
- 6: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT:\*
- 7: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT:\*
- 8: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT:\*
- 9: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT:\*
- 10: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT:\*
- 11: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT:\*
- 12: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT:\*
- 13: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT:\*
- 14: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT:\*
- 15: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT:\*
- 16: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT:\*
- 17: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT:\*
- 18: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:\*
- 19: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT:\*
- 20: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:\*
- 21: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:\*
- 22: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:\*
- 23: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT:\*
- 24: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT:\*
- 25: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2003.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed,

and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	281.5	24.5	2296	19	AAV20662	Barley D-hordein g
2	281.5	24.5	2380	22	AAF63380	Barley cDNA involv
3	281.5	24.5	2434	22	AAF63379	Barley cDNA involv
4	156	13.6	1302	24	AAI70852	Wheat glutenin gen
5	130.5	11.4	477	19	AAV20828	Nucleotide sequenc
6	129.5	11.3	109519	22	AAV20828	Micromonospora DNA
7	127.5	11.1	1089	23	ABL12131	Drosophila melanog
8	127.5	11.1	3089	23	ABL12130	Human cDNA sequenc
9	126.5	11.0	1847	22	AAH17970	Gene #2353 used to
10	124.5	10.8	2042	24	ABN95855	Human immunodefici
11	123.5	10.8	4350	22	AAV20828	Thermus thermophil
12	121.5	10.6	4347	21	AAA70724	Human protein PPI3
13	121	10.5	1029	25	ABZ22145	Human musculoskele
14	121	10.5	2986	22	ABZ80413	cDNA encoding nove
15	121	10.5	9968	22	AAH36353	Nucleotide sequenc
16	121	10.5	9968	25	ABX59341	Human cDNA differe
17	120.5	10.5	1806	22	AAH77885	Human immune/haema
18	120.5	10.5	1806	24	ABK84228	Murine Cdc25A comp
19	120.5	10.5	7954	22	AAK73498	Human reproductive
20	120.5	10.5	21429	22	AAV57362	Human testicular a
21	119.5	10.4	2794	22	AAV57362	Zea mays viviparou
22	119.5	10.4	2794	23	ABL98401	Human osteoblast d
23	119.5	10.4	2456	22	AAV57362	Human osteoblast d
24	119	10.4	2456	25	ABX13448	Human osteoblast d
25	119	10.4	92638	24	ABQ88096	Human osteoblast d
26	118.5	10.3	5822	20	AAV62933	Human osteoblast d
27	118.5	10.3	5822	20	AAV62933	Human osteoblast d
28	118.5	10.3	5822	20	AAV62933	Human osteoblast d
29	118.5	10.3	9615	22	AAH36590	Human musculoskele
30	118.5	10.3	9615	25	ABX59578	cDNA encoding nove
31	117.5	10.2	1966	20	AAK61220	Mouse DNA demethyl
32	117.5	10.2	5123	24	ABQ64378	Human von willebra
33	117	10.2	2429	23	ABL23142	Drosophila melanog
34	117	10.2	80557	24	ABX09142	Mycobacterium tube
35	117	10.2	4403765	22	AAI99683	Mycobacterium tube
36	117	10.2	4411529	22	AAI99682	Mycobacterium tube
37	116.5	10.1	3600	17	AAH13227	Thermotable enzym
38	116.5	10.1	5387	24	AAH39115	Human lung-specifi
39	116.5	10.1	10126	22	AAI99260	Human kidney relat
40	116.5	10.1	10126	22	AAI99260	Human kidney relat
41	116	10.1	707	22	AAH34931	Human colon cancer
42	116	10.1	1233	22	AAH44351	Human nucleotide s
43	116	10.1	1233	22	AAH44351	Heart muscle cell
44	116	10.1	1233	22	AAH49586	Human coding sequen
45	116	10.1	1630	22	AAH16370	Human cDNA sequenc

ALIGNMENTS

RESULT 1  
AAV20662  
ID AAV20662 standard; DNA; 2296 BP.  
XX  
AC AAV20662;  
XX  
DT 23-JUN-1998 (first entry)  
XX  
DE Barley D-hordein gene 5'-terminal region.  
XX  
KW Barley; D-hordein; DDP3; promoter; gene expression; regulation;  
XX seed; structural gene; db.  
OS Hordeum vulgare.  
XX  
PN WO9803655-A2.  
XX



PD 29-JAN-1998. 104  
 XX  
 XX  
 XX 22-JUL-1997; 97WO-JP02526. 366  
 XX  
 XX 23-JUL-1996; 96JP-0193433. 114  
 XX  
 XX (SAPB ) SAPPORO BREWERIES. 426  
 XX  
 PI Hirota N, Ito K, Kihara M, Kuroda H;  
 XX WPI; 1998-120779/11. 120  
 XX  
 XX Gene expression regulatory DNA, expression cassettes and vectors -  
 PT comprising promoter region from barley, Hordeum vulgare, D-hordein  
 PT gene, useful to control expression of desired gene e.g. to improve  
 PT seeds  
 PS  
 PS Example 3; Page 25-26; 42pp; English.  
 XX  
 XX The present sequence represents the 5'-terminal region of the barley  
 CC D-hordein gene. The present invention describes gene expression  
 CC regulatory DNA which comprises a promoter region derived from the  
 CC barley (Hordeum vulgare) D-hordein gene which allows expression of a  
 CC desired structural gene, and a regulatory region regulating such  
 CC expression. The introduction into plants of expression cassettes  
 CC containing the gene expression regulatory DNA (either directly or via  
 CC expression vectors) enables the expression of a gene within a plant cell  
 CC e.g. barley to be controlled. The use of activating and suppressing  
 CC regions in the regulatory DNA allows control of expression by e.g.  
 CC tissue type or developmental stage, whilst the use of only an activating  
 CC region maintains expression at a high level, providing an effective  
 CC production means when recovery of the product of the structural gene is  
 CC desired. The expression cassette/vector may be introduced into e.g.  
 CC maturing seed endosperm tissue or regeneratable plant cells (e.g. derived  
 CC from anthers) to improve seeds of barley or other plants, to produce  
 CC gene products in seeds or to contribute to plant breeding programmes.  
 CC The expression regulatory DNA can also be used in expression systems in  
 CC vitro. GUS activity in barley protoplasts transfected with plasmid  
 CC DP3GUS2 comprising isolated D-hordein promoter region was 1.5 times  
 CC higher than in those transfected with control pACT1f.  
 XX  
 SQ Sequence 2296 BP; 629 A; 708 C; 606 G; 353 T; 0 other;  
 Alignment Scores:  
 Pred. No.: 1.15e-08 Length: 2296  
 Score: 281.50 Matches: 79  
 Percent Similarity: 46.12% Conservative: 22  
 Best Local Similarity: 36.07% Mismatches: 51  
 Query Match: 24.52% Indels: 67  
 DB: 19 Gaps: 11

US-10-053-410-4 (1-206) x AAV20662 (1-2296)  
 QY 1 MetAlaLysIleAlaAlaAlaAlaAlaLeuCysPheAlaAlaLeuValAlaVal 20  
 DB 37 ATGGCTAAGCGGTGGTCTCTTTGGCGGTATC-----GTCGCCCTCGTGGCTTC 90  
 QY 21 AlaValCysGlnGlyGlyValGlu-----ArgGlnArgLeuArgAsp 34  
 DB 91 ACCACCGCTGAACGTGAGATCAATGGGAACAACATTTCCTGTAGCCGCTTAGGCAG 150  
 QY 35 LeuGlnCysTrpGlnGluValGlnGluSerProLeuAspAlaCysArgGlnValLeuAsp 54  
 DB 151 CTACAGTGTGAGCGGAGCTCCAGGAGAGCTCGCTGAGCGGTGCGCGGTGCGGAC 210  
 QY 55 ArgGlnLeuThrGlyGlyGlyGlyGlyGlyValGlyValGlyProPheArgTrpGlyThrGly 74  
 DB 211 CACAGCTG-----GTTGGCAGCTGCCATGACACGGGG 246  
 QY 75 LeuArgMetArgCysCysGlnGlnLeuGlnAspValSerArgGluCysArgCysAlaAla 94  
 DB 247 CTCAGATGACGTGCTGCCAGCAGCTTCGGGAGAGTTCAGCCCCGAGTGCAGCGCGCTGCC 306

QY 95 IleArgSerMetValArgGlyTyrGluGlu----- 104  
 DB 307 CTCAGCCAGGTGCTGAGGCAATACGAGCAGAAACGAGGTGCCATCCAGGAGGATCC 366  
 QY 105 -----AlaMetProProLeuGluLys---GlyTyrTrp 114  
 DB 367 TTCTACCGCGGGGACCGCAGCGCGCTGCAGCAAGGAGGATGGTGGGGAACCTCTGTA 426  
 QY 115 -----ProTyrGlyArgGlnGln----- 120  
 DB 427 AAATGGTACTACCCAGACCAAACTTCTTCGCAACAGTCAATGCAAGGGCAACAAGGGTAC 486  
 QY 121 -----GlnProProGlnGlyGlyGlyGlyGlyGlnGlyGly 133  
 DB 487 CACCAAGCGTAACCTTCTCCAGCAGCCAGGACGAGGAGGAGGAGGAGGAGGAGGAGG 546  
 QY 134 TyrTyrTyrProCysSerArgProGlyGluGlyTyrGlyTyrGlyGlnGlyGlnArg 153  
 DB 547 TCAACTTTCCCG---CAGCAGCGCAGGACAAGGACAA-----CAACGAGGACAGAGG 594  
 QY 154 Gln-----MetTyrPro-----ProCysArgProGlyThrThrGlyGly 166  
 DB 595 CAGCCATGCTCTATCCAAAGTCAACTTCCCAACAGCCAGGCGCAAGGGCAAGGG 651  
 RESULT 2  
 ID AAF63380 standard; cDNA; 2380 BP.  
 XX  
 AC AAF63380;  
 XX  
 DT 10-MAY-2001 (first entry)  
 XX  
 DE Barley cDNA involved in D-hordein production SEQ ID 2.  
 KW Barley; gel protein; D-hordein; malt; wort filterability; extraction; ss.  
 OS Hordeum vulgare.  
 XX  
 PN WO200111946-A1.  
 XX  
 PD 22-FEB-2001.  
 PF 16-AUG-2000; 2000WO-JP05476.  
 PR 16-AUG-1999; 99JP-0229696.  
 XX  
 PA (SAPB ) SAPPORO BREWERIES LTD.  
 XX  
 PI Hirota N, Kihara M, Ito K;  
 XX  
 DR WPI; 2001-191587/19.  
 XX  
 PT Transformation of barley with a D-hordein expression regulator for  
 PT production of barley with lowered gel protein content -  
 XX  
 PS Claim 4; Page 33-34; 40pp; Japanese.  
 XX  
 CC This invention relates to barley having a low gel protein content. The  
 CC barley is transformed with a polynucleotide sequence which regulated the  
 CC formation of D-hordein. Transformation results in the production of  
 CC barley strains with improved malting properties such as wort  
 CC filterability and efficiency of extraction. The present sequence  
 CC represents cDNA involved in the regulation of D-hordein production.  
 XX  
 SQ Sequence 2380 BP; 642 A; 747 C; 622 G; 369 T; 0 other;  
 Alignment Scores:  
 Pred. No.: 1.18e-08 Length: 2380  
 Score: 281.50 Matches: 79  
 Percent Similarity: 46.12% Conservative: 22  
 Best Local Similarity: 36.07% Mismatches: 51  
 Query Match: 24.52% Indels: 67  
 DB: 22 Gaps: 11

```

US-10-053-410-4 (1-206) x AAF63380 (1-2380)
QY 1 MetAlaLysIleAlaAlaAlaAlaAlaLeuCysPheAlaAlaLeuValAlaVal 20
DB 24 ATGGCTAAGCGGTGCTCTCTTTGGCGGTAATC-----GTGCCCTCGTGGCTCTC 77
QY 21 AlaValCysGlnGlyValGlu-----ArgGlnArgLeuArgAsp 34
DB 78 ACCACCGCTGAACGCTGAGATCAATGGGAACAACATTTCTTGTATAGCGCTCTAGGCAG 137
QY 35 LeuGlnCysTrpGlnGluValGlnGlnSerProLeuAspAlaCysArgGlnValLeuAsp 54
DB 138 CTACAGTGTGAGCGCGAGCTCCAGGAGAGCTCGCTCGAGCGCTGCGCGGGTCTGTGGAC 197
QY 55 ArgGlnLeuThrGlyGlyGlyGlyGlyValGlyValGlyProPheArgTrpGlyThrGly 74
DB 198 CAACAGCTG-----GTTGGCCAGCTGCCATGGAGCAGCGGG 233
QY 75 LeuArgMetArgCysGlnGlnLeuGlnAspValSerArgGluCysArgCysAlaAla 94
DB 234 CTCAGATGCACTGTCTGCCAGCAGCTTCGGGAGCTCAGCCCGAGTGCCTCGCTCGCC 293
QY 95 IleArgSerMetValArgGlyTyrGluGlu-----104
DB 294 CTCAGCGAGTGTGAGGCAATACGAGCAGCAAAACGAGGTGCCATCCAAAGGGAGGATCC 353
QY 105 -----AlaMetProProLeuGluLys---GlyTrpTrp-----114
DB 354 TTCTACCGCGGGGAGCCCGCCGCTGCAGCAAGGAGGATGGTGGGGAACCTCTGTGA 413
QY 115 -----ProTrpGlyArgGlnGln-----120
DB 414 AAATGGTACTACCAGACCAAACTTTCGCAACAGTCAATGCGAAGGGCAACAGGGTAC 473
QY 121 -----GlnProProProGlnGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGly 133
DB 474 CACCAAGAGGTAACTTCTTCCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 533
QY 134 TyrTyrTyrProCysSerArgProGlyGluGlyTyrGlyTyrGlyGlnGlyGlnArg 153
DB 534 TCAACTTTTCCCG---CAGCAGCAGCAGCAAGGACAA-----CAACAGGACAGAG 581
QY 154 Gln-----MetTyrPro-----ProCysArgProGlyThrThrGlyGly 166
DB 582 CAGCCATGGTCTCTATCCAGTCAACTTTCCCAACAGCAGCGGCAAGGGCAAGGG 638

RESULT 3
AAF63379
ID AAF63379 standard; cDNA; 2434 BP.
XX
AC AAF63379;
XX
DT 10-MAY-2001 (first entry)
XX
DE Barley cDNA involved in D-hordein production SEQ ID 1.
XX
KW Barley; gel protein; D-hordein; malt; wort filterability; extraction; ss.
XX
OS Hordeum vulgare.
XX
PN WO200111946-A1.
XX
PD 22-FEB-2001.
XX
PF 16-AUG-2000; 2000WO-JP05476.
XX
PR 16-AUG-1999; 99JP-0229696.
XX
PA (SABP) SAPPORO BREWERIES LTD.
XX
PI Hirota N, Kihara M, Ito K;
XX

```

```

DR WPI; 2001-191587/19.
XX Transformation of barley with a D-hordein expression regulator for
XX production of barley with lowered gel protein content
XX Claim 4; Page 31-33; 40pp; Japanese.
XX This invention relates to barley having a low gel protein content. The
XX barley is transformed with a polynucleotide sequence which regulated the
XX formation of D-hordein. Transformation results in the production of
XX barley strains with improved malting properties such as wort
XX filterability and efficiency of extraction. The present sequence
XX represents cDNA involved in the regulation of D-hordein production.
XX Sequence 2434 BP; 675 A; 757 C; 626 G; 376 T; 0 other;

Alignment Scores:
Pred. No.: 1,218-08 Length: 2434
Score: 281.50 Matches: 79
Percent Similarity: 46.12% Conservative: 22
Best Local Similarity: 36.07% Mismatches: 51
Query Match: 24.52% Indels: 67
DB: 22 Gaps: 11

US-10-053-410-4 (1-206) x AAF63379 (1-2434)
QY 1 MetAlaLysIleAlaAlaAlaAlaAlaLeuCysPheAlaAlaLeuValAlaVal 20
DB 37 ATGGCTAAGCGGTGCTCTCTTTGGCGGTAATC-----GTGCCCTCGTGGCTCTC 90
QY 21 AlaValCysGlnGlyValGlu-----ArgGlnArgLeuArgAsp 34
DB 91 ACCACCGCTGAACGCTGAGATCAATGGGAACAACATTTCTTGTATAGCGCTCTAGGCAG 150
QY 35 LeuGlnCysTrpGlnGluValGlnGlnSerProLeuAspAlaCysArgGlnValLeuAsp 54
DB 151 CTACAGTGTGAGCGCGAGCTCCAGGAGAGCTCGCTCGAGCGCTGCGCGGGTCTGTGGAC 210
QY 55 ArgGlnLeuThrGlyGlyGlyGlyGlyValGlyValGlyProPheArgTrpGlyThrGly 74
DB 211 CAACAGCTG-----GTTGGCCAGCTGCCATGGAGCAGCGGG 246
QY 75 LeuArgMetArgCysGlnGlnLeuGlnAspValSerArgGluCysArgCysAlaAla 94
DB 247 CTCAGATGCACTGTCTGCCAGCAGCTTCGGGAGCTCAGCCCGAGTGCCTCGCTCGCC 306
QY 95 IleArgSerMetValArgGlyTyrGluGlu-----104
DB 307 CTCAGCGAGTGTGAGGCAATACGAGCAGCAAAACGAGGTGCCATCCAAAGGGAGGATCC 366
QY 105 -----AlaMetProProLeuGluLys---GlyTrpTrp-----114
DB 367 TTCTACCGCGGGGAGCCCGCCGCTGCAGCAAGGAGGATGGTGGGGAACCTCTGTGA 426
QY 115 -----ProTrpGlyArgGlnGln-----120
DB 427 AAATGGTACTACCAGACCAAACTTTCGCAACAGTCAATGCGAAGGGCAACAGGGTAC 486
QY 121 -----GlnProProProGlnGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGly 133
DB 487 CACCAAGAGGTAACTTCTTCCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 546
QY 134 TyrTyrTyrProCysSerArgProGlyGluGlyTyrGlyTyrGlyGlnGlyGlnArg 153
DB 547 TCAACTTTTCCCG---CAGCAGCAGCAGCAAGGACAA-----CAACAGGACAGAG 594
QY 154 Gln-----MetTyrPro-----ProCysArgProGlyThrThrGlyGly 166
DB 595 CAGCCATGGTCTCTATCCAGTCAACTTTCCCAACAGCAGCGGCAAGGGCAAGGG 651

RESULT 4
AAI70852
ID AAI70852 standard; DNA; 1302 BP.

```

XX AC AAI70852;  
 XX AT 18-FEB-2002 (first entry)  
 XX DT  
 XX DE Wheat glutenin gene variant.  
 XX KW  
 XX KW wheat; glutenin; storage protein; variant; transgenic plant;  
 XX KW gluten; ds.  
 XX OS  
 XX OS Triticum aestivum.  
 XX FH Key Location/Qualifiers  
 XX FT allele replace(1181,C)  
 XX FT /\*tag= a  
 XX PN WO200179477-A2.  
 XX PN 25-OCT-2001.  
 XX PF 13-APR-2001; 2001WO-HU00045.  
 XX PR 14-APR-2000; 2000HU-0001563.  
 XX XX (MAGY ) MTA MEZOGRAZDASAGI KI.  
 XX PA Juhasz A, Tamas L, Bedo Z, Vida G, Karsai I, Lang L;  
 XX PI Tamasne Nyitrai E;  
 XX XX WPI; 2002-017612/02.  
 XX DR P-PSDB; AAM50386.  
 XX XX  
 XX PT A new nucleic acid comprising a variant allele of the 1Ax2asteriskHMW  
 XX PT glutenin subunit gene of wheat has at least one extra cysteine codon  
 XX PT and is useful for developing new wheat varieties with enhanced quality  
 XX PT gluten  
 XX PS Disclosure; Fig 4b; 28pp; English.  
 XX XX  
 XX CC The present sequence is that of the coding region of a newly  
 XX CC identified allele, designated 1Ax2asteriskB, of the 1Ax2asterisk  
 XX CC HMW glutenin subunit gene, found in certain lines of wheat  
 XX CC variety Bankuti 1201. The variant differs from the 1Ax2asterisk  
 XX CC gene by a single point mutation of C to G at nucleotide position  
 XX CC 1181. This results in a change from Ser to Cys at amino acid  
 XX CC position 394. Because of this, the newly identified allele  
 XX CC provides enhanced possibilities for the formation of disulfide  
 XX CC bridges compared with the known allele, and this could at least  
 XX CC partly explain the good technological properties of the variety.  
 XX CC The newly identified gene can be used to develop transgenic  
 XX CC wheat varieties with better quality gluten.  
 XX SQ Sequence 1302 BP; 439 A; 364 C; 327 G; 172 T; 0 other;

Alignment Scores:  
 Pred. No.: 0.313 Length: 1302  
 Score: 156.00 Matches: 62  
 Percent Similarity: 31.60% Conservative: 17  
 Best Local Similarity: 24.80% Mismatches: 43  
 Query Match: 13.59% Indels: 128  
 DB: 24 Gaps: 11

US-10-053-410-4 (1-206) x AAI70852 (1-1302)

QY 11 AlaLeuCysPheAlaLeuValAlaValCysGlnGlyGluValGluArgGln 30  
 |||||  
 DB 25 GCGGGGTAGTCGTCGCCCTTGTGGCTCTACCGCTCTGAAGGTAGGGCTCTGGGCAA 84  
 |||||  
 QY 31 ArgLeuArgAspLeuGlnCysTrpGlnGluValGlnGluSerProLeuAspAlaCysArg 50  
 |||||  
 DB 85 -----CTACAGTGTAGCGCGAGCTCCAGGAGCACTCGCTTAAGGCATGCCGA 132  
 |||||  
 QY 51 GlnValLeuAspArgGlnLeuThrGlyGlyGlyGlyGlyValGlyProPheArg 70  
 |||||

DB 133 CAGGTCTAGAC----- 144  
 |||||  
 QY 71 TrpGlyThrGlyLeuArgMetArgCysGlnGlnLeuGlnAspValSerArgGluCys 90  
 |||||  
 DB 145 -----CAGCAGCTCCGAGAGCTTAGCCCCGAGTGC 174  
 |||||  
 QY 91 ArgCysAlaAlaIleArgSerMetValArgGlyTrpGluGlu-----AlaMetProPro 108  
 |||||  
 DB 175 CAACCGTCGCGCGGCCGCGTCCGAGACAATATGACGACGAAGTCTGTTGGTGGCCGCC 234  
 |||||  
 QY 109 LeuGluLysGlyTrpTrpPro----- 115  
 |||||  
 DB 235 AAGGGTGGATCTTTTACCCCGCGAGACCACGCCACACAGCACTCCCAACAAGATATA 294  
 |||||  
 QY 116 ---TrpGly----- 117  
 |||||  
 DB 295 CTTTGGGGAATACCTGCACCTACTAAGAAGGTATTACCTAAGTGTAACTTCTCCGCAACAG 354  
 |||||  
 QY 118 -----ArgGlnGlnPro----- 122  
 |||||  
 DB 355 GTTTCATATATCCAGGCCAAGCTTCTTCGCAACGCGCCAGGACAAGGACAACAAGATATAC 414  
 |||||  
 QY 123 -----ProGlnGlnGly-----GlyGlyGlyGlnGlyGly 133  
 |||||  
 DB 415 TACCTAACTTCTCCGCAACAGTCAGGACATATGCAACACCGGACCAAGGGCAATCAGGG 474  
 |||||  
 QY 134 Tyr-----TyrTrpProCysSer 139  
 |||||  
 DB 475 TACTACCCAACTTCTCCGAGCAGTCAGGACAAAAGCAACACGAGGTACTATCCAACTTCT 534  
 |||||  
 QY 140 -----ArgProGly 142  
 |||||  
 DB 535 CCATGCGACGCCAGACAATGTGCAACCAACCAACAGGGGCAACAAGACAGCAACACCGA 594  
 |||||  
 QY 143 GluGlyTyrGlyTrpGlyGlnGly-----GlyGlnArgGln 154  
 |||||  
 DB 595 CAGGGCGCACTAAGACAGGACAACAAGGTCAGCAGTCAGGACCAAGGGCAACCAACCA 654  
 |||||

RESULT 5  
 AAV20828  
 ID AAV20828 standard; DNA; 477 BP.  
 XX  
 AC AAV20828;  
 XX  
 DT 20-JUL-1998 (first entry)  
 XX  
 DE Nucleotide sequence of Synthetic HMW-glutenin construct.  
 XX  
 KW HMW glutenin subunit repeat; wheat; synthetic; repetitive domain;  
 KW viscoelastic property; wheat flour; wheat dough; bread; noodle  
 KW ss.  
 XX  
 OS Synthetic.  
 OS Triticum sp.  
 XX  
 FH Key Location/Qualifiers  
 FT CDS 1..477  
 FT /\*tag= a  
 FT /product= "Synthetic HMW-glutenin protein"  
 XX  
 PN WO9808607-A1.  
 XX  
 PD 05-MAR-1998.  
 XX  
 XX 29-AUG-1997; 97WO-US15398.  
 XX  
 PR 30-AUG-1996; 96US-0706391.  
 XX

(USDA ) US SEC OF AGRIC.

PA  
XX  
PI  
XX  
DR  
XX  
DR  
XX  
XX  
PT  
PT  
PT  
XX  
PS  
XX  
XX  
CC  
CC  
CC  
CC  
CC  
CC  
XX  
SQ

Anderson OD;  
WPI: 1998-179218/16.  
F-PSDB; AAW53261.  
A process for altering the viscoelasticity of a dough - comprises genetic engineering of a high molecular weight glutenin subunit containing a non-natural repetitive domain  
Example 2; Page 17; 28pp; English.  
This nucleotide sequence is of the synthetic HMW-glutenin construct which is introduced into the DNA of the cereal to form a non-natural repetitive domain which changes the viscoelastic property of a dough.  
The domain comprises HMW glutenin subunit repeats which are synthetic in sequence or non-natural in number. The dough is useful for making a wheat flour, wheat dough, bread or noodle.  
Sequence 477 BP; 91 A; 135 C; 134 G; 117 T; 0 other;

Alignment Scores:  
Pred. No.: 4.43 Length: 477  
Score: 130.50 Matches: 55  
Percent Similarity: 37.62% Conservative: 21  
Best Local Similarity: 27.23% Mismatches: 53  
Query Match: 11.37% Indels: 73  
DB: 19 Gaps: 11

US-10-053-410-4 (1-206) x AAV20828 (1-477)

QY 24 GlnGlyGluValGluArgGlnArgLeuArgAspLeuGlnCys-----TrpGlnGlu 40  
DB 4 GAAGGTGAAGCGTCTGAACAG-----CTGCAGTGCATCCGCACTGCAGGAA 51  
QY 41 ValGlnGluSerProLeuAspAlaCysArgGlnValLeuAspArgGlnLeuThrGlyGly 60  
DB 52 CTGCAGGACGTGAACCTGAAGCTTCCAGCAGGTATGAC----- 93  
QY 61 GlyGlyGlyGlyValGlyProPheArgTrpGlyThrGlyLeuArgMetArgCysCys 80  
DB 93 ----- 93  
QY 81 GlnGlnLeuGlnAspValSerArgGluCysArgCysAlaAlaIleArgSerMetValArg 100  
DB 94 CAGCAGCTCGTGACATCTCTCCGGAATGCCACCCGGTGTGTTTCTCCGGTTGCTGCG 153  
QY 101 GlyTyrGluGlu-----AlaMetProProLeuGluGlyGlyTyrTrpProTrpGlyArg 118  
DB 154 CAGTACGACACAGCAGATCGTTGTTCGCGC-----AAAGTGGTACCTTCTATCCGGT 207  
QY 119 GlnGlnGlnProProGln-----GlyGlyGlyGlyGlyGlnGly 132  
DB 208 GAACACACTCCGCGCAGCAGCTGCAGCAGCTATCTCTGGGGCATCCCGGTCTGCTG 267  
QY 133 GlyTyrTyrTyrPro-----CysSerArgProGlyGluGlyTyrGlyGln 149  
DB 268 AGCGGTACTACCCGCTGTACTTCTGTCGCGAG----- 300  
QY 150 GlyGlyGlnArgGlnMetTyrPro-----ProCysArgProGlyThrThr--- 164  
DB 301 -----CAGGTTCTTACTACCCGGGTCCAGGTTCTCCGAGCGTCTAGTTCTTCTTAC 354  
QY 165 -----GlyGlyGlyProArgIleGlyArgValArgLeuThrIysAlaArgGluTyrAla 182  
DB 355 CAGGTTCTTCTGAACACACAGCGTGCATCTCTGAAGTTGCTAAAGCTCAGCAGCTGGCT 414  
QY 183 AlaGlyLeuProMetMetCysArgLeuSerGluProGlnGluCysSerIlePheSerGly 202  
DB 415 GCGCAGCTCCGGCTATGTGCGCGCTG-----GAAGGT 447  
QY 203 GlyAsp 204

Db 448 GGTGAC 453

RESULT 6  
AAS08693  
ID AAS08693 standard; DNA; 109519 BP.  
XX  
AC AAS08693;  
XX  
DT 26-SEP-2001 (first entry)  
XX  
DE Micromonospora DNA encoding biosynthetic enzymes for Everninomycin.  
XX  
KW Everninomycin; antibiotic; bottle-neck gene; orthomycin;  
XX fermentation; ds.  
OS Micromonospora carbonacea var. africana.  
XX  
FH Key Location/Qualifiers  
CDS complement (132..1382)  
FT /\*tag= a  
FT /product= "Evdh"  
FT complement (1389..1394)  
FT /\*tag= b  
FT complement (1490..2611)  
FT /\*tag= c  
FT /product= "Evdh"  
FT complement (2618..2622)  
FT /\*tag= d  
FT complement (2622..3860)  
FT /\*tag= e  
FT /product= "Evdh"  
FT complement (3867..3870)  
FT /\*tag= f  
FT 4143..5312  
FT /\*tag= g  
FT /product= "Evdh"  
FT 4134..4138  
FT /\*tag= h  
FT 5309..6235  
FT /\*tag= i  
FT /product= "Evdh"  
FT 6232..7275  
FT /\*tag= j  
FT /product= "Evdh"  
FT 6226..6229  
FT /\*tag= k  
FT 7272..8327  
FT /\*tag= l  
FT /product= "Evdh"  
FT 8342..9364  
FT /\*tag= m  
FT /product= "Evdh"  
FT 8333..8336  
FT /\*tag= n  
FT complement (9463..10224)  
FT /\*tag= o  
FT /product= "Evdh"  
FT complement (10232..10235)  
FT /\*tag= p  
FT 10424..11176  
FT /\*tag= q  
FT /product= "Evdh"  
FT 12027..12455  
FT /\*tag= r  
FT /product= "Evdh"  
FT /partial  
FT /note= "No start codon"  
FT complement (12108..13022)  
FT /\*tag= s  
FT /product= "Evdh"  
FT complement (13027..13030)  
FT /\*tag= t



```
QY 55 ArgGlnLeuThrGlyGlyGlyGlyGlyGlyValGlyProPheArgTrpGlyThrGly 74
DB 87615 ACGCGCGATCGACCGCTGGTGTGTGTCGCGGGCTGGTCCG----- 87656
QY 75 LeuArgMetArgCysCysGlnGlnLeuGlnAspValSerArgGluCysArgCysAlaAla 94
DB 87657 -----ACAAAGCTCCCCCAGGTGTACGGCG 87680
QY 95 IleArgSerMetValArgGlyTyrGluGluAlaMetPro---ProLeuGluLysGlyTrp 113
DB 87681 GTGCGAACCTGTGCGCGGGCGGTACTTCAACTGTCCGCGCCGAGCGACGGGGTGG 87740
QY 114 TrpProTrpGlyArgGlnGlnPro----- 122
DB 87741 TGGCGGTGTGCGCGCGCGCGCGCGCTGTGCGCTGTGTCAGCGTGATCGCCCGG 87800
QY 123 -----Pro-ProGlnGlyGlyGlyGlyGlyGlyGlyTyrTyrProCysSe 139
DB 87801 CGATCGTACACCGCAACACGGTGTGTGTCGCGCTTCGCGACCCACCGCCCTGGCTCGG 87860
QY 139 rArgProGlyGluGlyTyrGlyTyr-----GlyGlnGlyGlyGlnArgGlnMe 155
DB 87861 TGACCTTGGCGAGGTGTGCGCACCTCCGACCTGCGCGCGGGTGTCAACGTCTGA 87920
QY 155 tTyrProProCysArgProGlyThrThrGly----- 165
DB 87921 CCGGTGCGATACCGAGAGCGGTGCGACCTCGCGCGCACCTGGACGTCAACGGATCG 87980
QY 166 -----GlyGlyProArgIleGlyArgValArg 174
DB 87981 ACCTGACCGGGGTGGCGACGGGTCTGCTCGCCACCGA 88017
RESULT 7
ABLI12131
ID ABLI12131 standard; cDNA; 1089 BP.
XX
AC ABLI12131;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 30875.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical; gene; ss.
XX
OS Drosophila melanogaster.
XX
PN WO200171042-A2.
XX
PD 27-SEP-2001.
XX
PF 23-MAR-2001; 2001WO-US09231.
XX
PR 23-MAR-2000; 2000US-191637P.
PR 11-JUL-2000; 2000US-0614150.
XX
PA (PEKE ) PE CORP NY.
XX
PI Venter JC, Adams M, Li PWD, Myers EW;
XX
DR WPI; 2001-656860/75.
DR P-PSDB; ABB68028.
XX
PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
XX
PS Claim 1; SEQ ID NO 30875; 21pp + Sequence Listing; English.
XX
CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
```

```
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABLI16176-ABLI30511), expressed DNA
CC sequences (ABLI01840-ABLI16175) and the encoded proteins
CC (ABB57737-ABB72072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 1089 BP; 271 A; 334 C; 299 G; 185 T; 0 other;
Alignment Scores:
Pred. No.: 14.6 Length: 1089
Score: 127.50 Matches: 55
Percent Similarity: 34.4% Conservative: 16
Best Local Similarity: 26.70% Mismatches: 61
Query Match: 11.11% Indels: 74
DB: 23 Gaps: 11
US-10-053-410-4 (1-206) x ABLI12131 (1-1089)
QY 6 AlaAlaAlaAlaAlaAlaLeuCysPheAlaAlaLeu----- 17
DB 6 GCGATCTTCCGCGAGCTGTGTGTGTGTGTCGCGCGCGCTGTGCTGTGCTTTTGGCATCCAT 65
QY 18 -----ValAlaValAlaValCysGlnGlyGluValGluAlaArgGlnLeuArg 33
DB 66 CTGATACGACGTCGTCTCGCGCAATTTGCAGTCTCTCTCTCTCGCGCGGAGCTGT 125
QY 34 AspLeuGlnCysTrpGlnGluValGlnGluSerProLeuAspAlaCysArgGlnValLeu 53
DB 126 TGCCTCCGCGGTGGGTGCAC-----TGT----- 149
QY 54 AspArgGlnLeuThrGlyGlyGlyGlyGlyValGlyProPheArgTrp----- 71
DB 150 -----GGTGGCCGAGCTGCGCGCTTTCGGAGTTATCAATGGACTGAT 191
QY 72 -----GlyThrGlyLeuArgMetArgCysCys 80
DB 192 CAATGCCATGCACAAACACCGCTTCACACGAGCGCGAGTTCGAGAGACTGCAGCA 251
QY 81 GlnGlnLeuGlnAspValSerArgGluCysArgCysAlaAlaAlaArgSerMetValArg 100
DB 252 AGGAAGCGCGGGATGGGTCCGAATCGAGGAGGAATGCACAGACGAGGATTTGCACCACC 311
QY 101 -GlyTyrGluGluAlaMetProLeuGluLysGlyTrp-----TyrProTrp-- 116
DB 312 CAACCGTCAGCAACCGACGCTCCT-----GGCTGGCCACCGCGCTGGCAGTGGG 362
QY 117 -----GlyArgGlnGlnPro-----ProPr 124
DB 363 ACCCGTGGAAATCAGAAATCAACCGGATTGACTGGGTTCAACAGGATTTGCACCAGG 422
QY 124 oGlnGlyGlyGlyGlyGlyGlyGlyTyrTyrTyrProCysSerArgProGly----G1 143
DB 423 CTGGAATGGAGTGGCGGCGCAAGGA-----CCAGGATGGAA 458
QY 143 uGlyTyrGlyTyrGlyGlnGlyGlyGlnArgGlnMetTyrProProCysArgProGlyTh 163
DB 459 TGGTCCAGGATGGAACCGAGCGCGGTGAAGAGGA-----CCTCCACCAAGACCGGTTT 512
QY 163 rThrGlyGlyGlyPro 168
DB 513 TAAACGGCGCGGACCA 528
RESULT 8
ABLI12130
ID ABLI12130 standard; cDNA; 3089 BP.
XX
AC ABLI12130;
XX
DT 26-MAR-2002 (first entry)
XX
```





CC polynucleotide which comprises a 3'-end sequence, where the  
CC oligonucleotide comprises at least 15 nucleotides and the combination of  
CC the 5'-end sequence/3'-end sequence is selected from those defined in  
CC in gene therapy. The primers are useful for synthesising polynucleotides,  
CC particularly full-length cDNAs. The primers are also useful for the  
CC detection and/or diagnosis of the abnormality of the proteins encoded by  
CC the full-length cDNAs. The primers allow obtaining of the full-length  
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and  
CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to  
CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632  
CC represent oligonucleotides, all of which are used in the exemplification  
CC of the present invention.

XX Sequence 1847 BP; 467 A; 505 C; 431 G; 444 T; 0 other;

Alignment Scores:  
Pred. No.: 27.4 Length: 1847  
Score: 126.50 Matches: 62  
Percent Similarity: 42.93% Conservative: 17  
Best Local Similarity: 33.70% Mismatches: 60  
Query Match: 11.02% Indels: 45  
DB: 22 Gaps: 9

US-10-053-410-4 (1-206) x AAH17970 (1-1847)

QY 18 ValAlaValAlaValCysGlnGlyGluValGluArgGlnArgLeuArgAspLeuGlnCys 37  
DB 1550 GTGGAGTGGAGGATGCCAGGCGCTCCCGGAAATGGAGGAGGTGAGGGATTCAGGAC 1491  
QY 38 TrpGlnGluValGlnGlnSerProLeu-AspAlaCysArgGlnValLeuAspArgGlnLe 57  
DB 1490 CACCAGAGAGGGGAGGAGGAGGAGGTCCTCC-----TTCTCCAGGCAAGG 1440  
QY 57 uThr-GlyGlyGlyGly-----GlyGlyGly-ValGlyProPhe 69  
DB 1439 GAGGAGTGGGGGGGAATTCACGACTCCAGGCAAGAGGAGTGGTGGGGGATTCT 1380  
QY 70 -----ArgTrpGlyThrGlyLeuArgMetArgCysGlnGln 82  
DB 1379 AGCACTCCAGGCAAGAGGAGGAGTGGGGGATGGCAGTACTCCAGGCAAGAGAGGG 1320  
QY 83 LeuGlnAspValSer-----ArgGluCysArgCysAlaAlaIleArgSer 97  
DB 1319 TGAAGGATGCCAACACCTCAGGCAAGAGGAGGGGTGGAGGATGTAGCATCCCGACA 1260  
QY 98 MetValArgGlyTyr-GluGluAlaMetProProLeuGluLysGlyTrpTrpProTrpG 117  
DB 1259 CAAAGGAGGGGTGGAGAGATAGCATGACTCTCCAGGTAAAGAGGG-----GG 1212  
QY 117 YargGlnGlnProProPro-----GlnGlyGlyGlyGlyGlnGlyTyrTyrTy 136  
DB 1211 TGAGGAGATGCAACACCCCGAGCAAGAGAGGTGGAGGAGGAGGAGGA----- 1160  
QY 136 rProCysSerArgProGlyGlyGlyTyrGlyTyr-----GlyGlnGlyGly 152  
DB 1159 -----GGAGGAGGAGGAGGAGTGGTACTATCCCGAGGAGGAGTGGTGG 1116  
QY 152 nArgGlnMetTyrPro-----ProCysArgProGlyThrThrGlyGlyProArg 170  
DB 1115 TGAATAATAGTCCAGAGTCCACAGGTAAAGAGGAGGGGAGGAGGAGGAGGAGGAGGAG 1056  
QY 170 eGly 171  
DB 1055 AGCA 1052

RESULT 10

ABN95855/C

ID ABN95855 standard; DNA; 2042 BP.

XX

AC

XX

DT 13-AUG-2002 (first entry)

XX Gene #2353 used to diagnose liver cancer.

XX Gene; liver cancer; ds; hepatocellular carcinoma; hepatotropic;  
KW metastatic liver tumour; cytostatic; expression profile; disease state;  
KW disease progression; drug toxicity; drug efficacy; drug metabolism.  
XX Homo sapiens.

XX WO200229103-A2.

XX 11-APR-2002.

XX 02-OCT-2001; 2001WO-US30589.

XX 02-OCT-2000; 2000US-237054P.

XX (GENE-) GENE LOGIC INC.

XX Horne D, Alvares C, Peres-Da-Silva S, Vockley JG;

XX WPI; 2002-426119/45.

XX Diagnosing and detecting the progression of liver cancer,  
PT hepatocellular carcinoma or metastatic liver tumor in a patient,  
PT involves detecting the level of expression of two or more genes in a  
PT liver tissue sample -  
XX Claim 1; SEQ ID NO 2353; 298pp; English.

XX The invention relates to a novel method for diagnosing and detecting the  
CC progression of liver cancer, hepatocellular carcinoma or metastatic liver  
CC tumor in a patient, and differentiating metastatic liver cancer from  
CC hepatocellular carcinoma in a patient, involving detecting the level of  
CC expression of two or more genes represented in ABN93503-ABN97455 in a  
CC tissue sample. The method of the invention has hepatotropic, and  
CC cytostatic activity. The method is useful for diagnosing and detecting  
CC the progression of liver cancer, hepatocellular carcinoma and metastatic  
CC liver carcinoma in a patient. The method is useful for identifying  
CC expression profiles which serve as useful diagnostic markers as well as  
CC markers that can be used to monitor disease states, disease progression,  
CC drug toxicity, drug efficacy and drug metabolism.  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.

XX Sequence 2042 BP; 326 A; 755 C; 611 G; 350 T; 0 other;

Alignment Scores:

Pred. No.: 39.9 Length: 2042  
Score: 124.50 Matches: 54  
Percent Similarity: 31.84% Conservative: 10  
Best Local Similarity: 26.87% Mismatches: 47  
Query Match: 10.84% Indels: 90  
DB: 24 Gaps: 11

US-10-053-410-4 (1-206) x ABN95855 (1-2042)

QY 48 AlaCysArgGlnValLeuAspArgGlnLeuThrGlyGly-GlyGlyGlyGlyValG 67  
DB 899 GCGTCCCCGAGTCCGCCCGCCCG-----GGAGGAAGGAGGAGGCGGTGGGTCCG 849  
QY 67 YProPheArgTrpGlyThrGlyLeuArgMetArgCysGlnGlnLeuGlnAspVal 87  
DB 848 C-----AGGTGGGGT-----GGGGT----- 833  
QY 87 rArgGluCysArgCysAlaAlaIleArgSerMetValArgGlyTyrGluGluAlaMetPr 107  
DB 832 -----TGGCGGAGGAGGAGAAACCCACACCCCTGTC 804  
QY 107 oProLeuGluLysGlyTrpTrpProTrpGlyArgGlnGlnProPro----- 123  
DB 803 TCCTCTCAGCAACTCCCGGGCGCCCTGGGGGCAACTCCCGTCCCGCCACCGCCCATGGA 744





```
QY 83 LeuGlnAspValSerArgGluCysArgCysAlaAlaIleArgSerMetValArgGlyTyr 102
Db 528 -----CCGCTACGGGACGAGATCGTGGCCCT 554
QY 103 GluGluAlaMetProProLeuGluLysGlyTyrTrpProTrpGlyArgGlnGlnPro 122
Db 555 GAGCGCGGCTTCCTCGGCGCTTCGCCCC---CATCTCGGGAGGTCCACGCGCCCT 611
QY 123 ProProGlnGlyGly-----GlyGlyGlyGlnGlyGlyTyrTyrProCys 138
Db 612 CGCGCGCAAGAGCGGGGCTTCGCTTGGAGGAGACCGCGGGGAGAGGGGTGCTCGGGC 671
QY 139 SerArgProGlyGluGlyTyrGlyTyrGlyGlnGlyGlyArgGlnMetTyrProPro 158
Db 672 CCTCGAGCCAGCGGCGGAGGCGGGAACGGGGCCAGACCTGGTGGGGCCACCG 731
QY 159 CysArgProGlyThrThrGlyGlyGly-----ProArgIle 170
Db 732 GGACGACCTGCTCTCTCTCGAGGGCGCGCCGCCACCGGTTCGCCAGCGCGGGGA 791
QY 171 GlyArg 172
Db 792 GGCCAA 797
RESULT 14
ID ABZ80413/c
XX ABZ80413 standard; cDNA; 2986 BP.
AC ABZ80413;
XX
XX 28-MAY-2003 (first entry)
XX
XX Human protein PP13671 encoding cDNA SEQ ID NO:18.
XX
XX Human; cancer; cancer suppression; cancer inhibitor; gene; ss.
XX
XX Homo sapiens.
XX
XX Key Location/Qualifiers
XX CDS 1314..2414
XX FT /*tag= a
XX FT /product= "Human protein PP13671"
XX
XX CN1368509-A.
XX
XX PD 11-SEP-2002.
XX
XX PF 08-FEB-2001; 2001CN-0105310.
XX
XX PR 08-FEB-2001; 2001CN-0105310.
XX
XX (SHAN-) SHANGHAI INST ONCOLOGY.
XX
XX Gu J;
XX
XX WPI: 2003-112778/11.
XX
XX P-PSDB; ABP96556.
XX
XX Human protein that suppresses cancer cell growth and its coding
XX sequence -
XX
XX Claim 5; Page 24-26 (Disclosure); 36pp; Chinese.
XX
XX ABZ80408 to ABZ80418 encode the human proteins ABP96551 to ABP96561 which
XX have cancer inhibiting functions. Also described is a method for
XX preparing the proteins using recombination techniques. The human proteins
XX from the present invention, and nucleotide sequences encoding them, can
XX be used for treating diseases such as cancer.
XX
XX Sequence 2986 BP; 581 A; 976 C; 942 G; 487 T; 0 other;
SQ
Alignment Scores:
```

```
Pred. No.: 92.9 Length: 2986
Score: 121.00 Matches: 59
Percent Similarity: 30.47% Conservative: 12
Best Local Similarity: 25.32% Mismatches: 62
Query Match: 10.54% Indels: 100
DB: 25 Gaps: 10
US-10-053-410-4 (1-206) x ABZ80413 (1-2986)
QY 5 AlaAlaAlaAlaAlaAlaLeuCysPheAlaAlaLeuValAlaValAlaValCysGln 24
Db 1292 GCCCTGCCGGCACCCCTCTCTGAGCTGGCCCTCCCTCGTC----- 1251
QY 25 GlyGluValGluArgGlnArgLeuArgAspLeuGlnCysTrpGlnGlnValGlnGlnSer 44
Db 1250 -----CTCCCTCCCTGGGGCGAGACGACGACGAGTCG 1221
QY 45 ProLeuAspAlaCysArgGlnValLeuAspArgGlnLeuThrGlyGlyGlyGlyGly 64
Db 1220 CCCTCCACCGTCTGTACTCCAGCCACC----- 1194
QY 65 GlyValGlyProPheArgTrpGlyThrGlyLeuArgMetArgCys-CysGlnGln----- 82
Db 1193 ---CTGCCTCTGTGAGATGGGA-----CGCTGCATGCTTAGAGGTCC 1152
QY 83 -----LeuGlnAspValSerArgGluCysArgCys-AlaAlaIleA 96
Db 1151 CTGTTCTCCCTACCCCTTTGCTGGGTCTCCCGGCCCCACAGATGACGACCTGTC 1092
QY 96 rGserMetValArgGlyTyrGluGluAlaMetProProLeuGluLysGlyTyrTrpPro 116
Db 1091 GATATAAGCTC-----TGGCCCT 1074
QY 116 rPglyArgGlnGlnProProGlnGlyGly----- 127
Db 1073 GGCGCGGCCAGCTCGAGGACCGGCCAGTCTGGATGTTTCAGACCTGCGGGGCCCT 1014
QY 128 -----GlyGlyGlyG 131
Db 1013 GGGGCTCTCCCTAGGACGAGGGCCACCTCTCTGGGGTTTCTTAGTCAGGACAA 954
QY 131 InGlyGlyTyrTyrTyrProCysSerArgProGlyGluGlyTyrGlyTyrGlyGly 151
Db 953 CAGGGGGCTCTGTAGAGCCA-----CGCGTGCAGGGGGCTGGGTACGGGTACA----- 904
QY 151 lYGlnArgGlnMetTyrProProCysArgProGlyThrThrGlyGlyGlyProArgIle 171
Db 903 -----CCTCCCTCGCCCTGGGACACCGAGTCTCTCCGATACTTGGG 861
QY 171 lYArgValArgLeuThrLysAlaArgGluTyrAlaAlaGlyLeuProMetCysArgL 191
Db 860 GTCCGGCAGCT-----GCAGGGGATATCCACCCCTGCA----- 826
QY 191 euSerGluProGlnGluCysSerIlePheSerGly 202
Db 825 --GAGGAGCCCGAGGAGCTCGGCGTCCGCGCTGGC 793
RESULT 15
AAL36353
ID AAL36353 standard; DNA; 9968 BP.
XX
XX AAL36353;
XX
XX 08-JAN-2002 (first entry)
XX
XX Human musculoskeletal system related polynucleotide SEQ ID NO 2718.
XX
XX Cytostatic; immunosuppressive; nootropic; neuroprotective; antiviral;
XX anti-allergic; hepatotropic; antidiabetic; antiinflammatory; antitumor;
XX vulnary; anticonvulsant; antibacterial; antifungal; antiparasitic;
XX cardiant; gene therapy; cancer; immune disorder; cardiovascular disorder;
XX neurological disease; infection; human; secreted protein;
XX musculoskeletal system; ds.
```





GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: November 29, 2003, 18:16:16 ; Search time 311 Seconds  
(without alignments)  
2182.335 Million cell updates/sec

Title: US-10-053-410-4

Perfect score: 1148

Sequence: 1 MAKIAAAAAAALCFALNAV.....MMCRILSEPQECIFSGGDY 206

Scoring table:

BLOSUM62  
Xgapop 10.0, Xgapext 0.5  
Ygapop 10.0, Ygapext 0.5  
Fgapop 6.0, Fgapext 7.0  
Delop 6.0, Delext 7.0

Searched: 2190069 seqs, 1647345023 residues

Total number of hits satisfying chosen parameters: 4380138

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n model -DEV=xlh  
-Q/cgn2\_1/USFTO\_spool/US10053410/runat\_28112003\_140759\_27990/app\_query.fasta\_1.391  
-DB=Published Applications NA -QMT=fastap -SUFFIX=p2n.rnpb -MINMATCH=0.1  
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62  
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100  
-THR MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0  
-MAXLEN=200000000 -USER=US10053410 @CGN\_1\_1 271 @runat\_28112003\_140759\_27990  
-NCPUS=6 -ICPU=3 -NO MMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -DSFBLCK=100  
-LONGLOG -DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5  
-FGAPOP=6 -FGAPEXT=7 -XGAPOP=10 -XGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published Applications NA:

1: /cgn2\_6/prodata/2/pubpna/US07\_PUBCOMB.seq:\*  
2: /cgn2\_6/prodata/2/pubpna/PCT\_NEW\_PUB.seq:\*  
3: /cgn2\_6/prodata/2/pubpna/US06\_NEW\_PUB.seq:\*  
4: /cgn2\_6/prodata/2/pubpna/US06\_PUBCOMB.seq:\*  
5: /cgn2\_6/prodata/2/pubpna/US07\_NEW\_PUB.seq:\*  
6: /cgn2\_6/prodata/2/pubpna/PCTUS\_PUBCOMB.seq:\*  
7: /cgn2\_6/prodata/2/pubpna/US08\_NEW\_PUB.seq:\*  
8: /cgn2\_6/prodata/2/pubpna/US08\_PUBCOMB.seq:\*  
9: /cgn2\_6/prodata/2/pubpna/US09A\_PUBCOMB.seq:\*  
10: /cgn2\_6/prodata/2/pubpna/US09B\_PUBCOMB.seq:\*  
11: /cgn2\_6/prodata/2/pubpna/US09C\_PUBCOMB.seq:\*  
12: /cgn2\_6/prodata/2/pubpna/US09\_NEW\_PUB.seq:\*  
13: /cgn2\_6/prodata/2/pubpna/US10A\_PUBCOMB.seq:\*  
14: /cgn2\_6/prodata/2/pubpna/US10B\_PUBCOMB.seq:\*  
15: /cgn2\_6/prodata/2/pubpna/US10\_NEW\_PUB.seq:\*  
16: /cgn2\_6/prodata/2/pubpna/US60\_NEW\_PUB.seq:\*  
17: /cgn2\_6/prodata/2/pubpna/US60\_PUBCOMB.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	127.5	11.1	9025608	14	US-10-156-761-1

C	2	123.5	10.8	4350	14	US-10-168-843A-33	Sequence 33, Appli
	3	123	10.7	747	12	US-10-259-165-748	Sequence 748, App
	4	122	10.6	930	14	US-10-156-761-4504	Sequence 4504, Ap
	5	122	10.6	1578	14	US-10-156-761-3369	Sequence 3369, Ap
	6	122	10.6	9025608	14	US-10-156-761-1	Sequence 1, Appli
	7	121	10.5	1029	11	US-09-938-901-5	Sequence 5, Appli
	8	121	10.5	9968	10	US-09-764-877-2718	Sequence 2718, Ap
C	9	120.5	10.5	1806	12	US-10-168-097A-1	Sequence 1, Appli
	10	119.5	10.4	2794	11	US-09-764-891-8525	Sequence 8525, Ap
C	11	118.5	10.3	700	12	US-10-027-632-140092	Sequence 140092,
	12	118.5	10.3	700	13	US-10-027-632-140092	Sequence 140092,
C	13	118.5	10.3	2988	12	US-10-027-632-111991	Sequence 111991,
	14	118.5	10.3	2988	13	US-10-027-632-111991	Sequence 111991,
C	15	118.5	10.3	3031	12	US-10-027-632-115601	Sequence 115601,
	16	118.5	10.3	3031	13	US-10-027-632-115601	Sequence 115601,
C	17	118.5	10.3	9615	10	US-09-764-877-2955	Sequence 2955, Ap
	18	118	10.3	5862	12	US-10-132-134-15	Sequence 15, Appli
C	19	118	10.3	52101	12	US-10-132-134-1	Sequence 1, Appli
C	20	117.5	10.2	582	12	US-10-027-632-140201	Sequence 140201,
	21	117.5	10.2	582	13	US-10-027-632-140201	Sequence 140201,
C	22	117.5	10.2	583	12	US-10-027-632-16086	Sequence 16086, A
	23	117.5	10.2	583	13	US-10-027-632-16086	Sequence 16086, A
C	24	117.5	10.2	699	12	US-10-027-632-105402	Sequence 105402,
	25	117.5	10.2	699	13	US-10-027-632-105402	Sequence 105402,
C	26	117	10.2	80557	15	US-10-080-170-647	Sequence 647, App
	27	116.5	10.1	5387	13	US-10-001-873-22	Sequence 22, Appli
C	28	116	10.1	707	14	US-10-106-698-2023	Sequence 2023, Ap
	29	116	10.1	1233	10	US-09-749-728B-2	Sequence 2, Appli
C	30	116	10.1	1329	14	US-10-156-761-6978	Sequence 6978, Ap
	31	115.5	10.1	7727	12	US-10-204-925-5	Sequence 5, Appli
C	32	115	10.0	2152	12	US-09-873-319-712	Sequence 712, App
	33	115	10.0	2152	12	US-09-960-706-1073	Sequence 1073, Ap
C	34	115	10.0	2432	9	US-09-925-301-574	Sequence 574, App
C	35	115	10.0	2432	12	US-10-262-473-5	Sequence 5, Appli
C	36	115	10.0	3148	14	US-10-084-817-219	Sequence 219, App
	37	114.5	10.0	1299	14	US-10-156-761-4590	Sequence 4590, App
	38	114	9.9	4167	9	US-09-764-878-282	Sequence 282, App
	39	114	9.9	4167	9	US-09-764-860-1145	Sequence 1145, Ap
C	40	114	9.9	4167	11	US-09-764-846-345	Sequence 345, App
	41	114	9.9	4167	11	US-09-764-891-7710	Sequence 7710, Ap
C	42	114	9.9	4167	11	US-09-764-891-10231	Sequence 10231, A
	43	114	9.9	4167	12	US-10-212-872-1145	Sequence 1145, Ap
C	44	114	9.9	4167	14	US-10-091-483-345	Sequence 345, App
	45	114	9.9	4167	14	US-10-079-854-282	Sequence 282, App

#### ALIGNMENTS

#### RESULT 1

US-10-156-761-1/c  
; Sequence 1, Application US/10156761  
; Publication No. US20030119018A1  
; GENERAL INFORMATION:  
; APPLICANT: OMURA, SATOSHI  
; APPLICANT: IKEDA, HARUO  
; APPLICANT: ISHIKAWA, JUN  
; APPLICANT: HORIKAWA, HIROSHI  
; APPLICANT: SHIBATA, TADAYOSHI  
; APPLICANT: SAKAKI, YOSHIYUKI  
; APPLICANT: HATTORI, MASAHIRA  
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
; FILE REFERENCE: 249-262  
; CURRENT APPLICATION NUMBER: US/10/156, 761  
; PRIOR FILING DATE: 2002-05-29  
; PRIOR APPLICATION NUMBER: JP 2001-204089  
; PRIOR FILING DATE: 2001-05-30  
; PRIOR APPLICATION NUMBER: JP 2001-272697  
; PRIOR FILING DATE: 2001-08-02  
; NUMBER OF SEQ ID NOS: 15109  
; SEQ ID NO 1  
; LENGTH: 9025608  
; TYPE: DNA  
; ORGANISM: Streptomyces avermitilis

```
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (4187715)
; OTHER INFORMATION: a, t, c, g, other or unknown
US-10-156-761-1
Alignment Scores:
Pred. No.: 14.1 Length: 9025608
Score: 127.50 Matches: 42
Percent Similarity: 40.88% Conservative: 14
Best Local Similarity: 30.66% Mismatches: 45
Query Match: 11.11% Indels: 36
DB: 14 Gaps: 7

US-10-053-410-4 (1-206) x US-10-156-761-1 (1-9025608)
QY 59 GlyGlyGlyGlyGlyGlyValGlyValGlyThrGlyLeuArgMetArg 78
|||
|||
|||
Db 4272268 GGTTCGAACCTGGGGCGGCATATGACCGGGCGCTGGCCACCGGGACTCCGCGCGGC 4272209
|||
|||
|||
QY 79 CysCysGlnLeuGlnAspValSerArgGluCys-----ArgCysAlaAlaIleArg 96
|||
|||
|||
Db 4272208 TGCTGCGCGCTTTCGCG-----TGCTGGCGCGGTGCTGCTGAGGCTGCG 4272164
|||
|||
|||
QY 97 SerMetValArgGlyTyrgluGluAlaMetProLeuGluLysGlyTTP----- 113
|||
|||
|||
Db 4272163 GGCCTGGGGCTGGGTGG-----CGGGGGTGGTGGCGGTTTG 4272128
|||
|||
|||
QY 114 -----TTPProTTPGlyArgGlnGlnProProGlnGly-Gly-- 127
|||
|||
|||
Db 4272127 CGCGGGGGTGACGTGGGTGGTTCACGCGGGAGCGCGTGGCGGAGGCGC 4272068
|||
|||
|||
QY 128 -----GlyGlyGlnGlyGlyTyrglyTyrglyProCysSerAr 140
|||
|||
|||
Db 4272067 GCTGTGCGGGCGCGACAGCGCGGTGGCGGGTGGCGAGGTGC-----TGTGCGGG 4272011
|||
|||
|||
QY 140 gProGlyGluGlyTyrglyTyrglyGlnGlyGly-----GlnArgGlnMetTyPr 157
|||
|||
|||
Db 4272010 GCCGTTCCTGGATATGCGACGCGTGGCGGGCGCGCTCAGGGGTGGCACGCTGG 4271951
|||
|||
|||
QY 157 oProCysArgProGlyThrThrGlyGlyProArgIleGlyArgVal 173
|||
|||
|||
Db 4271950 CCCTGGTTCGGCGGTGGGTGGCGAGTGGCGGTGCCACGCTGGGTG 4271902
|||
|||
|||
RESULT 2
US-10-168-843A-33/c
; Sequence 833, Application US/10168843A
; Publication No. US20030108562A1
; GENERAL INFORMATION:
; APPLICANT: Medical Research Council
; APPLICANT: International Aids Vaccine Initiative
; APPLICANT: University of Nairobi
; TITLE OF INVENTION: Improvements in or Relating to Immune Responses to HIV
; FILE REFERENCE: MJL/C1248/1/M
; CURRENT APPLICATION NUMBER: US/10/168,843A
; CURRENT FILING DATE: 2002-09-24
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 33
; LENGTH: 4350
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Chimeric
; OTHER INFORMATION: polynucleotide
US-10-168-843A-33
Alignment Scores:
Pred. No.: 0.0229 Length: 4350
Score: 123.50 Matches: 42
Percent Similarity: 38.89% Conservative: 7
Best Local Similarity: 33.33% Mismatches: 54
Query Match: 10.76% Indels: 23

; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (4187715)
; OTHER INFORMATION: a, t, c, g, other or unknown
US-10-053-410-4 (1-206) x US-10-168-843A-33 (1-4350)
QY 55 ArgGlnLeuThrGlyGlyGlyGly-----GlyGlyGlyValGlyProPheArgTrpGly 72
|||
|||
|||
Db 2296 CGGCGCGCCAGCGCGCGCGGCTTCTTGGCGGCGCTGGCGGAGCCCTTGGGCTTGGGT 2237
|||
|||
|||
QY 73 ThrGlyLeuArgMetArg-----CysCysGlnGlnLeuGlnAspValSerArg 88
|||
|||
|||
Db 2236 AGGGGTTCGACGTGTACAGATCTTTGATGATCGGATGCGCTTCAGCAGCTCTCTCGTCGG 2177
|||
|||
|||
QY 89 GluCysArgCysAlaAlaIleArgSerMetValArgGlyTyrgluGluAlaMetProPro 108
|||
|||
|||
Db 2176 AGTTGCGCGAGCGGC-----CGGCGCTCTCGGGGTGGTCTCGGTCTTGGACTCCA 2126
|||
|||
|||
QY 109 LeuGluLysGlyTTPTrpProTTPGlyArgGlnGlnProProGlnGlyGlyGly 128
|||
|||
|||
Db 2125 ---CCTTCTTCTTGGACTCTTGGGCGCGTGGAGATGCCCT-----GGGTCTGGG 2078
|||
|||
|||
QY 129 Gly-----GlyGlnGlyGlyTyrglyTyrglyProCysSerArgProGly 142
|||
|||
|||
Db 2077 GATGGGCTGCTTGGGAGTGGGTCTTGGTGT-----CCTTGTGGACTGGGGG 2027
|||
|||
|||
QY 143 ---GluGlyTyrglyTyrglyGlnGlyGlnArgGlnMetTyrglyProCysArgPro 161
|||
|||
|||
Db 2026 TGCGCGCGCGCTGGCGCGCTTCTTGGCGCGCTAGGAGATGCCAGGCCCTTGTTCAGGA 1967
|||
|||
|||
QY 162 GlyThrGlyGlyGly 167
|||
|||
|||
Db 1966 AGCACACGCGGCGAGTGGT 1949
|||
|||
|||
RESULT 3
US-10-259-165-748
; Sequence 748, Application US/10259165
; Publication No. US20030135888A1
; GENERAL INFORMATION:
; APPLICANT: Zhu, Tong
; APPLICANT: Chang, Hur-song
; APPLICANT: Briggs, Steven P.
; APPLICANT: Cooper, Bret
; APPLICANT: Glazebrook, Jane
; APPLICANT: Goff, Stephen A.
; APPLICANT: Katagiri, Fumiyaki
; APPLICANT: Kreps, Joel
; APPLICANT: Moughamer, Todd
; APPLICANT: Provart, Nicholas
; APPLICANT: Riche, Darrell
; TITLE OF INVENTION: GENES THAT ARE MODULATED BY POSTTRANSCRIPTIONAL GENE SILENCING
; FILE REFERENCE: 70030-NP
; CURRENT APPLICATION NUMBER: US/10/259,165
; CURRENT FILING DATE: 2002-09-26
; PRIOR APPLICATION NUMBER: US 60/370,620
; PRIOR FILING DATE: 2002-04-04
; PRIOR APPLICATION NUMBER: US 60/368,327
; PRIOR FILING DATE: 2002-03-27
; PRIOR APPLICATION NUMBER: US 60/325,277
; PRIOR FILING DATE: 2001-09-26
; NUMBER OF SEQ ID NOS: 782
; SOFTWARE: PatentList.pl version 3.0.4 (C) 2001 Syngenta
; SEQ ID NO 748
; LENGTH: 747
; TYPE: DNA
; ORGANISM: Zea mays
US-10-259-165-748
Alignment Scores:
Pred. No.: 0.00475 Length: 747
Score: 123.00 Matches: 43
Percent Similarity: 31.45% Conservative: 7
Best Local Similarity: 27.04% Mismatches: 47
Query Match: 10.71% Indels: 62
```



```
DB: 12 Gaps: 6
US-10-053-410-4 (1-206) x US-10-259-165-748 (1-747)
QY 59 GlyGlyGlyGlyGlyGlyValGlyProPhe----- 69
DB 241 GCGGAGGCGGAGGAGCGCGGCTACGGCTCCCGCGCGGTGCGGATCTGGC 300
QY 70 -----ArgTTPGlyThrGlyLeuArgMet----- 77
DB 301 GCGCGGCTGCAGCTACGCTGTGTAGTGGGCGCGCGGAGATCCGCGCGCGGGGC 360
QY 78 -----ArgCys-----CysGlnGlnLeuGlnAspValSerArgGluCysA-gCys 92
DB 361 GGTCCCGCGCGTGTACAAAGTGGCGAGCGCGCGGCACATGTCGAAGGACTGC----- 414
QY 93 AlaAlaIleArgSerMetValArgGlyTyrGluGluAlaMetProProLeuGluLysGly 112
DB 415 -----CCTAGCGCGCGAGCGCGGA 432
QY 113 TrpTrpProTTPGlyArgGlnGlnProProProGlnGlyGlyGlyGlnGly 132
DB 433 GCGCGGTACGCGGAGCGGCTACGCGCGCGCGGTACGAGAGAGAGCGCGCGCGGC 492
QY 133 GlyTyrTyrTyrProCysSerArgProGlyGlu----- 143
DB 493 GGTGGCTGTTCAGTGTGGCGAGCTGGCCACATGGCCAGGAGTGTCTCCAGCGCGGC 552
QY 144 ---GlyTyrGlyTyrGlyGlnGlyGly-----Gln 152
DB 553 GCGCGGTACGCGGTGGCGCGCGCGGTGGAGCGCGGTGTACAACTGCGCGCGAGGC 612
QY 153 ArgGlnMetTyrProProCysA-gProGlyThrThrGlyGlyGlyProArgIleGly 171
DB 613 GGCCACATGGCCAGGAGTGCCTCCAGCGGTGGCGCGGTGGCGAGGAGTTCGGC 669

RESULT 4
US-10-156-761-4504
; Sequence 4504, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; PRIOR FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 4504
; LENGTH: 930
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(930)
US-10-156-761-4504
Alignment Scores:
Pred. No.: 0.00729 Length: 930
Score: 122.00 Matches: 50
Percent Similarity: 36.49% Conservative: 4
Best Local Similarity: 33.78% Mismatches: 45
Query Match: 10.63% Indels: 50
DB: 14 Gaps: 8
```

```
US-10-053-410-4 (1-206) x US-10-156-761-4504 (1-930)
QY 59 GlyGlyGlyGlyGlyGlyValGlyProPheArgTTP-----Gly 72
DB 480 GCGGAGGCGGAGGCGGAGCGCGGACCGCGCGGTTCGAGGTGACGTCCGATCCGGA 539
QY 73 ThrGlyLeuArgMetArgCysGlnGlnLeuGlnAspValSerArgGluCysA-gCys 92
DB 540 GCGGTGCGCGAAGCGGAGCTTGCGC-----GCAAGCGAAGCGGAGCTGC 587
QY 93 AlaAlaIleArgSer-----MetValArgGlyTyrGluGlu 104
DB 588 GCGCAAGCGGAGCGGAGCGCGCGCTCCCGCGTGTCTCCCGGAGGATTCGACGA 647
QY 105 AlaMetProProLeuGluLysGly-----TTP 114
DB 648 GGAGGAGACCTCTGTCGCGGAGGCTTCTGTGGCGAGGGGAGCGCCACCGATGCT- 706
QY 115 ProTTPGlyArgGlnGlnProPro-----Pr 124
DB 707 TCGTGGGAGAGCGGCGCGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 766
QY 124 oGlnGlyGlyGlyGlyGlyGlnGlyGlyTyrTyrTyrProCysSerArgProGluGlu 144
DB 767 GCAGCGCGGAGGAGGTGGACCGCGCGCGCGCGCGGTTA---CGCCCGGTCCCG 823
QY 144 YTyrGlyTyrGlyGlnGlyGlyGlnArgGlnMetTyrProProCysArgProGlyThr 164
DB 824 CGCGGA-----GAGACACCGCGGAGCGGACCTCCGCTG----- 860
QY 164 rGlyGlyGlyProArgIleGly 171
DB 861 ----GGCGGCGCGCGCGGCGGC 878

RESULT 5
US-10-156-761-3369
; Sequence 3369, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; PRIOR FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 3369
; LENGTH: 1578
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1578)
US-10-156-761-3369
Alignment Scores:
Pred. No.: 0.0121 Length: 1578
Score: 122.00 Matches: 63
Percent Similarity: 32.76% Conservative: 13
Best Local Similarity: 27.16% Mismatches: 65
Query Match: 10.63% Indels: 92
DB: 14 Gaps: 12
```



Alignment Scores:  
Pred. No.: 0.01 Length: 1029  
Score: 121.00 Matches: 41  
Percent Similarity: 39.34% Conservative: 7  
Best Local Similarity: 33.61% Mismatches: 44  
Query Match: 10.54% Indels: 30  
DB: 11 Gaps: 4

US-10-053-410-4 (1-206) x US-09-938-901-5 (1-1029)

QY 63 GlyGlyValGlyProPheArgTrpGlyThrGlyLeuArgMetArgCysGlnGln 82  
DB 486 GCGCGCGGGAGGGCTTTCGCCCTGGGACCGGAGCTGCG 527  
QY 83 LeuGlnAspValSerArgGluCysArgCysAlaAlaIleArgSerMetValArgGlyTyr 102  
DB 528 -----CCGCTACGGGAGCAGATCGTGGCCCT 554  
QY 103 GluGlnAlaMetProLeuGluLysGlyTyrTrpTrpProTrpGlyArgGlnGlnPro 122  
DB 555 GAGGCGCGCTTCTCGCGGCTTCGCCCC- -CATCTCGGGAGGTCCACGCGCCCT 611  
QY 123 ProProGlnGlyGly-----GlyGlyGlnGlyGlyTyrTyrProCys 138  
DB 612 CGCGCGCAAGGAGCGGGCTTCGCTTGGAGGAGACCGCGGGGAAGGGTGTCTCGGGC 671  
QY 139 SerArgProGlyGluGlyTyrGlyTyrGlyGlnGlyGlyGlnArgGlnMetTyrProPro 158  
DB 672 CTTGAGGCGCAGCGGGCGAGAGCGGGAACGCGGCGCAGACCTGTGGGGCCCCACCG 731  
QY 159 CysArgProGlyThrThrGlyGlyGly-----ProArgIle 170  
DB 732 GAGGAGCTGTCTTCTCTCTGAGGGGGCGCGCCGCCACCGTTCGCGAGCGCGGGGA 791  
QY 171 GlyArg 172  
DB 792 GGCCTAA 797

RESULT 8  
US-09-764-877-2718  
; Sequence 2718, Application US/09764877  
; Patent No. US20020147140A1  
; GENERAL INFORMATION:  
; APPLICANT: Rosen et al.  
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies  
; FILE REFERENCE: PC005  
; CURRENT APPLICATION NUMBER: US/09/764,877  
; Prior application data removed - refer to PALM or file wrapper  
; NUMBER OF SEQ ID NOS: 4031  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 2718  
; LENGTH: 9968  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-764-877-2718

Alignment Scores:  
Pred. No.: 0.0876 Length: 9968  
Score: 121.00 Matches: 60  
Percent Similarity: 37.57% Conservative: 11  
Best Local Similarity: 31.75% Mismatches: 65  
Query Match: 10.54% Indels: 54  
DB: 10 Gaps: 10

US-10-053-410-4 (1-206) x US-09-764-877-2718 (1-9968)

QY 5 AlaAlaAlaAlaAlaAlaLeuCysPheAlaAlaLeuValAlaValCysGln 24  
DB 89 GCGGCGCGCGGTGATACCGCGTGTCTCGCGCTCGCGCTCGCGCGCGCGCGCGCGCGCG 148  
QY 25 GlyGluValGluArgGlnArgLeuArgAspLeuGlnCysTrpGlnGlnValGlnGlnSer 44

DB 149 GGCACATCGACTTCAGGAGCTGCGC- -ATGATGCTGCTAAGCGAGCGCGCGCGCGCG 205  
QY 45 ProLeuAspAlaCysArgGlnValLeuAspArgGlnLeuThrGlyGlyGlyGlyGly 64  
DB 206 -----GAGAGACGTTCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 232  
QY 65 GlyValGlyProPheArg-----TrpGlyThrGlyLeuArgMetArgCysCys 80  
DB 233 GCGAG-GGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 273  
QY 81 GlnGlnLeuGlnAspValSerArgGluCysArgCysAlaAlaIleArgSerMetValArg 100  
DB 274 -----CGCGCTCGCGCTCGCGCTCGCGCTCGCGCTCGCGCTCGCGCTCGCGCTCGCG 312  
QY 101 GlyTyrGluGluAla-----MetProPro-LeuGluLysGlyTyr----- 113  
DB 313 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 372  
QY 114 -----TrpProTrpGlyArgGlnGlnProProProGlnGlnGlyGlyGly 128  
DB 373 CGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 432  
QY 129 -----GlyGlyGlnGly-----GlyTyrTyrTyrTyrTyrTyrTyrTyrTyrTyrTyr 141  
DB 433 GTGCGTGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 483  
QY 141 GglyGluGlyTyrGlyTyrGlyGlnGlyGlyGlnArgGlnMetTyrProPro---CysAr 160  
DB 484 GGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 543  
QY 160 gProGlyThrThrGlyGlyGlyPro 168  
DB 544 TCCGCGCGCACAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 589

RESULT 9  
US-10-168-097A-1/c  
; Sequence 1, Application US/10168097A  
; Publication No. US20030166245A1  
; GENERAL INFORMATION:  
; APPLICANT: CENTRE NATIONAL OF THE RECHERCHE SCIENTIFIQUE  
; APPLICANT: INSTITUT CURIE  
; TITLE OF INVENTION: WASP FAMILY PROTEIN FRAGMENTS, AND USES THEREOF  
; FILE REFERENCE: IFB99WASP  
; CURRENT APPLICATION NUMBER: US/10/168,097A  
; CURRENT FILING DATE: 2002-06-17  
; NUMBER OF SEQ ID NOS: 84  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 1  
; LENGTH: 1806  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (35)..(1540)  
US-10-168-097A-1

Alignment Scores:  
Pred. No.: 0.0191 Length: 1806  
Score: 120.50 Matches: 39  
Percent Similarity: 35.14% Conservative: 0  
Best Local Similarity: 35.14% Mismatches: 22  
Query Match: 10.50% Indels: 51  
DB: 12 Gaps: 5

US-10-053-410-4 (1-206) x US-10-168-097A-1 (1-1806)

QY 60 GlyGlyGlyGlyGlyGlyValGlyProPheArgTrpGlyThrGlyLeuArgMetArgCys 79  
DB 1284 GAGGAGGTGGGGAGGGCGTGTTCATTC- -GGAGCTGGCGCGCGCGCGCGCGCGCG 1232  
QY 80 CysGlnGlnLeuGlnAspValSerArgGluCysArgCysAlaAlaIleArgSerMetVal 99

```
Db 1232 ----- 1233
Qy 100 ArgGlyTyrGluGluAlaMetProProLeuGluLysGlyTrpTrpProTyrP-clyArgGl 119
Db 1231 -----TGGTGGCGGTGGTGGCAT 1211
Qy 119 nGlnGlnProProGlnGlnGlyGlyGlyGlnGlyGlyTyrTyrProCys 139
Db 1210 GGTGGCCACAGCTCCAGGGGGTGGAGGGGAGTGTCCA-----GA 1166
Qy 139 rArgProGlyGluGlyTyrGlyTyrGlyGlnGlyGlnArgGlnMetTyrProProCys 159
Db 1165 ACCTCCAGTAGCTGGAGGGGGTGGTGGAGGG-----CCCCC 1127
Qy 159 sArgProGlyThrThrGlyGlyGlyProArg 169
Db 1126 TCGGCCTGGG-----GGTGGGGTCCCCGG 1102

RESULT 10
US-09-764-891-8525
; Sequence 8525, Application US/09764891
; Publication No. US2003007808A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC006
; CURRENT APPLICATION NUMBER: US/09/764,891
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 10231
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 8525
; LENGTH: 2794
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-764-891-8525

Alignment Scores:
Pred. No.: 0.0361 Length: 2794
Score: 119.50 Matches: 40
Percent Similarity: 38.76% Conservatives: 10
Best Local Similarity: 31.01% Mismatches: 49
Query Match: 10.41% Indels: 30
DB: 11 Gaps: 6

US-10-053-410-4 (1-206) x US-09-764-891-8525 (1-2794)
Qy 60 GlyGlyGlyGlyGlyValGlyProPheArgTrpGlyThrGlyLeuArgMetArgCys 79
Db 103 GGTGGGGTGGGGGAGCTGGCCCTTGTGTGGCCCTCACCTTTCTCCACACA 162
Qy 80 CysGlnGlnLeuGlnAap-----ValSerArgGluCysArgCysAlaAla 95
Db 163 GGCCAGAACTACATGAAGGGCTCCCGGAATTGGAGAAGAGATTTCCTC----- 216
Qy 96 ArgSerMetValArgGlyTyrGluGluAlaMetProProLeuGluLysGlyTrpTrpPro 115
Db 217 -----TATCTGACCAATGAAGCCCTCTGGTACGGCTGGTAGTG 258
Qy 116 TrpGlyArgGlnGlnProProGlnGlyGlyGlyGlyGlnGly----- 132
Db 259 TGGGG-----CACCGGTGCTGGAGTGGGGCTCCCGGTCTGGCCTG 306
Qy 133 -----GlyTyrTyrProCysSerArgProGlyGlyGlyTyrGlyTyrGlyGlnGly-- 150
Db 307 ATGCACCTTCCCGCAGCTGTGAACCTCTCGAGAAGATGCTGGTGTGGACGCGAGC 366
Qy 151 -----GlyClnArgGlnMetTyrPro-----ProCysArgProGl 162
Db 367 AGCGGGTGACGGAGGCGCGCTGGCCCATCTTCTACTTCTGAGTCCCTCGACACACGG 426
Qy 162 yThrThrGlyGlyGlyProArgile 170
|||||
```

```
Db 427 AAGATGAGCCCGGAGGTCCAGAGTA 451

RESULT 11
US-10-027-632-140092/c
; Sequence 140092, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 140092
; LENGTH: 700
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-140092

Alignment Scores:
Pred. No.: 0.012 Length: 700
Score: 118.50 Matches: 54
Percent Similarity: 37.78% Conservatives: 14
Best Local Similarity: 30.00% Mismatches: 49
Query Match: 10.32% Indels: 63
DB: 12 Gaps: 10

US-10-053-410-4 (1-206) x US-10-027-632-140092 (1-700)
Qy 58 ThrGlyGlyGlyGlyGly----- 64
Db 504 ACCGGCGGGCGGAGCGCGCTCGGGGGGAGCTCGGCGAGCGCGCACGTCA 445
Qy 65 GlyValGlyProPheArgTrpGlyThrGlyLeuArgMetArgCysCysGln-GlnLeuGl 84
Db 444 GGGGCGGGCGCGCGCGCGGAAAGT-----CTCTGTAGGGCGCGCGGC 400
Qy 84 nAapValSerArgGluCys-----ArgCysAlaAla-1leArgSerMetValA 100
Db 399 TACCTCTACCGCGCGCGCGCTATGGCCCTTTCGGTCCCGGCTACTCACCGGGCTTCCGA 340
Qy 100 rGlyTyrGluGluAlaMetProProLeuGluLysGlyTrpTrpProTyrGlyArgGlnG 120
Db 339 AAGCGCGCGGAGTAGTGGCTCCGACGAAAGGCGCGGAGCGGTGGAGCTGGCGCC 280
Qy 120 lGlnProProGlnGlyGlyGlyGlyGlnGlyGlyTyrTyrProCysSerA 140
Db 279 TCCCCCGCGGTGAGCTGACGAGCGCGCGG-----CCCGCGCGCC 238
Qy 140 rProGlyGlyGlyTyrGlyTyrGlyGln-----GlyGlyGlnArgGln----- 154
Db 237 GCCTGTGGGGGGTGGCTCTTCGCCCTTTCCTGTGGGGGGGAGAGCGGTGGC 178
Qy 155 -----MetTyrProProCysArgProGlyT 163
Db 177 AGCGGGCGCGCGCGCGCTGCTCGAGGAGAACCCCTTCGCGCGCTGGACACCG---A 121
```

```
QY 163 hrThrGly-----GlyGlyProArgIleGlyArgValArgLeuThrLysAlaArgGluT 181
Db 120 CCGCGGGTCCCGCGGAGCCCCCGA-----CGGCGCGGCGCG 82
QY 181 yrAlaalaGlyLeuProMetCysArgLeuSerGluProGlnGluCysSerile 199
Db 81 GAGCAGCGGAGGCGCC-----GGTCCCGGTGAGTGTCTTGTA 44

RESULT 12
US-10-027-632-140092/c
; Sequence 140092, Application US/10027632
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 140092
; LENGTH: 700
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-140092

Alignment Scores:
Pred. No.: 0.012 Length: 700
Score: 118.50 Matches: 54
Percent Similarity: 37.78% Conservative: 14
Best Local Similarity: 30.00% Mismatches: 49
Query Match: 10.32% Indels: 63
DB: 13 Gaps: 10

US-10-053-410-4 (1-206) x US-10-027-632-140092 (1-700)
QY 58 ThrGlyGlyGlyGlyGly-----64
Db 504 ACCGCGGGGCGGAGCGCGCGTCCGGGGCGGAGCTCGGCGAGCGCCACGTCA 445
QY 65 GlyValGlyProPheArgTrpGlyThrGlyLeuArgMetArgCysGln-GlnLeuG1 84
Db 444 GGCGCGCGCGCGCGCGCGCGGAGT-----CTCTGTAGGCGGCGCGCG 400
QY 84 nAspValSerArgGluCys-----ArgCysAlaAla-IleArgSerMetValA 100
Db 399 TACCTTCAGCGCGCGCGCGCTATGCGCTTTCGTCGCGCGCTACTCAGCGGCTCCGA 340
QY 100 rGlyTyrGluGluAlaMetProProLeuGluLysGlyTyrTrpProTyrGlyArgGlnG 120
Db 339 AAGCCCGCCGAGTAGTGGCGGTCCGACGAAAGGCGCGGAGCGGTGAGCTGCCGCG 280
QY 120 lnglnProProPheArgTrpGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGly 140
Db 279 TCCCCCGCCGTCAGTGTACGAGCGCGCGCG-----CGCCGAGCG 238
QY 140 rGProGlyGluGlyTyrGlyGlyGln-----GlyGlyGlnArgGln----- 154
Db 237 GCCCTGGTGGCGGGTGCCTCTTCGCCCTTTCCTGCTCGCGGGGCGAGAGCGGTGCG 178
```

```
QY 155 -----MetTyrProProCysArgProGlyT 163
Db 177 AGCGCGCGCGCGCGCGCTGCTCGGAGAACCCCTTCGCCCGCTCGACACCG--A 121
QY 163 hrThrGly-----GlyGlyProArgIleGlyArgValArgLeuThrLysAlaArgGluT 181
Db 120 CCGCGGGTCCCGCGGAGCCCCCGA-----CGGCGCGGCGCG 82
QY 181 yrAlaalaGlyLeuProMetCysArgLeuSerGluProGlnGluCysSerile 199
Db 81 GAGCAGCGGAGGCGCC-----GGTCCCGGTGAGTGTCTTGTA 44

RESULT 13
US-10-027-632-111991/c
; Sequence 111991, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 111991
; LENGTH: 2988
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-111991

Alignment Scores:
Pred. No.: 0.0478 Length: 2988
Score: 118.50 Matches: 57
Percent Similarity: 33.85% Conservative: 8
Best Local Similarity: 29.69% Mismatches: 45
Query Match: 10.32% Indels: 82
DB: 12 Gaps: 12

US-10-053-410-4 (1-206) x US-10-027-632-111991 (1-2988)
QY 44 SerProLeuAspAlaCysArgGlnValLeuAspArgGlnLeuThrGlyGlyGlyGly 63
Db 2874 AGTCCCTCTCAGCC-----CAGTCCCGAGTGGGAGGCTCAG 2836
QY 64 GlyGlyValGlyProPheArgTrpGlyThrGlyLeuArgMetArgCysGlnGlnLeu 83
Db 2835 GCGGTCTCGGA-----GATACGGCC 2815
QY 84 GlnAspValSerArgGluCys-----ArgCysAla-----AlaIleArgSerMet 98
Db 2814 CAGGAGGAGAGCTCGCTCTGACCGCGCTGTGCCGAGAGCTCGGCATCGACAAGCG 2755
QY 99 ValArg-----GlyTyrGluGluAlaMetProProLeu----- 109
Db 2754 GTGAGGGGAGAGGTGGCCATCAAGAAAGCTGAGCCGACCTTTTCACTCCGAGATCTTCGC 2695
QY 110 -----GluLysGlyTyr 113
```

Db 2694 CAAGCGCGCTACCGGAGCTGCTGCTGAAGCACATGCAGCATAGTAGTGG 2635  
|||:|:|  
Qy 114 TrpProTrpGlyArgGlnGlnGlnProProGlnGlyGlyGlyGlyGlyGly 133  
|||:|:|  
Db 2634 TGG-----CTCCCGGGGAGGGGTGGGGCTCCCTGGG 2599  
|||:|:|  
Qy 134 TyrTyrProCysSerArgProGlyGluGlyTyrGly-----TyrGlyGln 149  
|||:|:|  
Db 2598 -----GAGTCAGGGGAGGGCGAGCGCGCGCTCCCGGTATGAGAG 2557  
|||:|:|  
Qy 150 -GlyGlyGlnArgGlnMetTyrProProCysArgProGlyThrThrGlyGlyProAr 169  
|||:|:|  
Db 2556 CTCGGACAGTCT-----CTCCCTGCTCCCTGAC----- 2526  
|||:|:|  
Qy 169 gileGlyArgValArgLeuThrLysAlaArgGluTyrAlaAlaGlyLeuProMetMetCy 189  
|||:|:|  
Db 2525 -----GGTGCCTCTCGCTCACTATTGAAGAGGGGGTGGCGAGACCTTCCC-----TG 2476  
|||:|:|  
Qy 189 sArgLeuSerGluProGlnGluCysSerIlePhe 200  
|||:|:|  
Db 2475 CCACATTTCCAGGAAGAGGTCAAAAGGTATTTTT 2442  
|||:|:|

## RESULT 14

US-10-027-632-111991/c  
; Sequence 111991, Application US/10027632  
; GENERAL INFORMATION:  
; APPLICANT: Wang, David G.  
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide  
; Polymorphisms in the Human Genome  
; FILE REFERENCE: 108827.129  
; CURRENT APPLICATION NUMBER: US/10/027,632  
; CURRENT FILING DATE: 2002-04-30  
; PRIOR APPLICATION NUMBER: US 60/218,006  
; PRIOR FILING DATE: 2000-07-12  
; PRIOR APPLICATION NUMBER: US 60/198,676  
; PRIOR FILING DATE: 2000-04-20  
; PRIOR APPLICATION NUMBER: US 60/193,483  
; PRIOR FILING DATE: 2000-03-29  
; PRIOR APPLICATION NUMBER: US 60/185,218  
; PRIOR FILING DATE: 2000-02-24  
; PRIOR APPLICATION NUMBER: US 60/167,363  
; PRIOR FILING DATE: 1999-11-23  
; PRIOR APPLICATION NUMBER: US 60/156,358  
; PRIOR FILING DATE: 1999-09-28  
; PRIOR APPLICATION NUMBER: US 60/146,002  
; PRIOR FILING DATE: 1999-08-09  
; NUMBER OF SEQ ID NOS: 325720  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 111991  
; LENGTH: 2988  
; TYPE: DNA  
; ORGANISM: Human  
US-10-027-632-111991

Alignment Scores:  
Pred. No.: 0.0478 Length: 2988  
Score: 118.50 Matches: 57  
Percent Similarity: 33.85% Conservatives: 8  
Best Local Similarity: 29.69% Mismatches: 45  
Query Match: 10.32% Indels: 82  
DB: 13 Gaps: 12

US-10-053-410-4 (1-206) x US-10-027-632-111991 (1-2988)

Qy 44 SerProLeuAspAlaCysArgGlnValLeuAspArgGlnLeuThrGlyGlyGlyGly 63  
|||:|:|  
Db 2874 AGTCCCTCTCAGCC-----CAGCTGCCAGTGGAGGGGTCTAG 2836  
|||:|:|  
Qy 64 GlyGlyValGlyProPheArgTrpGlyThrGlyLeuArgMetArgCysGlnGlnLeu 83  
|||:|:|  
Db 2835 GGCGGTCTGGGA-----GATACGGCC 2815  
|||:|:|

Qy 84 GlnAspValSerArgGluCys-----ArgCysAla-----AlaIleArgSerMet 98  
|||:|:|  
Db 2814 CAGGAGGAGAGCTCGCTGCTGCGCGCTGTGCCGACAGCTCGCCATCGACAAGCG 2755  
|||:|:|  
Qy 99 ValArg-----GlyTyrGluGluAlaMetProProLeu----- 109  
|||:|:|  
Db 2754 GTCAGGGGAGAGGTGGCCATCAAGAAAGCTGAGCCGACCCCTTTTTCAGTCCGAGATCTTCGC 2695  
|||:|:|  
Qy 110 -----GlyTyrGlyGlyTyr 113  
|||:|:|  
Db 2694 CAAGCGCGCTACCGGAGCTGCTGCTGAAGCACATGCAGCATGAGAAAGCTAGGTGG 2635  
|||:|:|  
Qy 114 TrpProTrpGlyArgGlnGlnGlnProProGlnGlyGlyGlyGlyGlyGly 133  
|||:|:|  
Db 2634 TGG-----CTGCCCCGGGAGGGGTGGGGCTCCCTGGG 2599  
|||:|:|  
Qy 134 TyrTyrProCysSerArgProGlyGluGlyTyrGly-----TyrGlyGln 149  
|||:|:|  
Db 2598 -----GAGTCAGGGGAGGGCGAGCGCGCGCTCCCGGTATGAGAG 2557  
|||:|:|  
Qy 150 -GlyGlyGlnArgGlnMetTyrProProCysArgProGlyThrThrGlyGlyProAr 169  
|||:|:|  
Db 2556 CTCGGACAGTCT-----CTCCCTGCTCCCTGAC----- 2526  
|||:|:|  
Qy 169 gileGlyArgValArgLeuThrLysAlaArgGluTyrAlaAlaGlyLeuProMetMetCy 189  
|||:|:|  
Db 2525 -----GGTGCCTCTCGCTCACTATTGAAGAGGGGGTGGCGAGACCTTCCC-----TG 2476  
|||:|:|  
Qy 189 sArgLeuSerGluProGlnGluCysSerIlePhe 200  
|||:|:|  
Db 2475 CCACATTTCCAGGAAGAGGTCAAAAGGTATTTTT 2442  
|||:|:|

## RESULT 15

US-10-027-632-1115601/c  
; Sequence 115601, Application US/10027632  
; Publication No. US20030204075A9  
; GENERAL INFORMATION:  
; APPLICANT: Wang, David G.  
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide  
; Polymorphisms in the Human Genome  
; FILE REFERENCE: 108827.129  
; CURRENT APPLICATION NUMBER: US/10/027,632  
; CURRENT FILING DATE: 2002-04-30  
; PRIOR APPLICATION NUMBER: US 60/218,006  
; PRIOR FILING DATE: 2000-07-12  
; PRIOR APPLICATION NUMBER: US 60/198,676  
; PRIOR FILING DATE: 2000-04-20  
; PRIOR APPLICATION NUMBER: US 60/193,483  
; PRIOR FILING DATE: 2000-03-29  
; PRIOR APPLICATION NUMBER: US 60/185,218  
; PRIOR FILING DATE: 2000-02-24  
; PRIOR APPLICATION NUMBER: US 60/167,363  
; PRIOR FILING DATE: 1999-11-23  
; PRIOR APPLICATION NUMBER: US 60/156,358  
; PRIOR FILING DATE: 1999-09-28  
; PRIOR APPLICATION NUMBER: US 60/146,002  
; PRIOR FILING DATE: 1999-08-09  
; NUMBER OF SEQ ID NOS: 325720  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 115601  
; LENGTH: 3031  
; TYPE: DNA  
; ORGANISM: Human  
US-10-027-632-115601

Alignment Scores:  
Pred. No.: 0.0485 Length: 3031  
Score: 118.50 Matches: 57  
Percent Similarity: 33.85% Conservatives: 8  
Best Local Similarity: 29.69% Mismatches: 45  
Query Match: 10.32% Indels: 82  
DB: 12 Gaps: 12

US-10-053-410-4 (1-206) x US-10-027-632-115601 (1-3031)

```
QY 44 SerProLeuAspAlaCysArgGlnValLeuAspArgGlnLeuThrGlyGlyGlyGly 63
Db 2904 AGTCCCTCTCAGCC-----CAGCTGCCAGTGGAGGGGTCTAG 2866
QY 64 GlyValGlyProPheArgTrpGlyThrGlyLeuArgMetArgCysGlnGlnLeu 83
Db 2865 GCGCGTCTGGGA-----GATACGGCC 2845
QY 84 GlnAspValSerArgGluCys-----ArgCysAla-----AlaIleArgSerMet 98
Db 2844 CAGGAGGAGAGCTCCCTCTGCTGACCGGCTGTGCCGACAGCTCGGCCATCGACAAGCG 2785
QY 99 ValArg-----GlyTyrGluGluAlaMetProProLeu----- 109
Db 2784 GTCAGGGGAGAGGTGGCCATCAGAGAGCTGAGCCGACCCCTTTCAGTCCGAGATCTTCGC 2725
QY 110 -----GluLysGlyTyr 113
Db 2724 CAAGCGCGCTACCGGGAGCTGCTGCTGAAGCACATGCAGCATGAGAACGTAGGTGG 2665
QY 114 TrpProTrpGlyArgGlnGlnGlnProProProGlnGlyGlyGlyGlnGlyGly 133
Db 2664 TGG-----CTGCCCCGGGAGAGGGGTGGGGCTGCCCTGGG 2629
QY 134 TyrTyrTyrProCysSerArgProGlyGluGlyTyrGly-----TyrGlyGln 149
Db 2628 -----CAGTCAGGGGAGGGCGGCGGCGGGGCTCCCGGTATGGAGAG 2587
QY 150 -GlyGlyGlnArgGlnMetTyrProProCysArgProGlyThrThrGlyGlyProAr 169
Db 2586 CTCGGGACAGTCT-----CCTCCTGCTCCCTGAC----- 2556
QY 169 gIleGlyArgValArgLeuThrLysAlaArgGluTyrAlaAlaGlyLeuProMetMetCy 189
Db 2555 -----GGTGCCTCTCGCCTCACTATTGAAAGAGGGGGTGGCGAGACCTTCCC-----TG 2506
QY 189 sArgLeuSerGluProGlnGluCysSerIlePhe 200
Db 2505 CCACATTCAGGAAGAGGTCAAGGGTATTTT 2472
```

Search completed: November 29, 2003, 21:28:17  
Job time : 6283 secs

**THIS PAGE BLANK (USPTO)**